START

0 009979

USEPA / Office of Toxic Substances

Graphical Exposure Modeling System

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

G E M S USER'S GUIDE

JUL 1990
RECEIVED
EDMC

General Sciences Corporation
March 1989

(GSC-TR-32-88-018)

This GEMS User's Guide was prepared for:
U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF PESTICIDES AND TOXIC SUBSTANCES
EXPOSURE EVALUATION DIVISION

Prepared under:

Task Nos. 1-36, 2-15

Contract No. 68-02-4281

Project Officer: Russell Kinerson

Task Managers: Patricia Harrigan
Cathy Turner

Prepared by:

Herb Hunt, Dalroy Ward, and Jon Chen

General Sciences Corporation
6100 Chevy Chase Drive
Laurel, Maryland 20707

DISCLAIMER

This edition of the user's guide is a preliminary document and is presented only as a DRAFT. It has not been released formally by the Office of Toxic Substances, Office of Pesticides and Toxic Substances, U.S. Environmental Protection Agency, and should not, at this stage of development, be construed to represent EPA policy. It is being circulated for comments on technical merit and policy implications.

Tradenames of commercial products may be included as an integral part of the documentation of this technical system. While trademarks or tradenames of commercial products may be included herein, such inclusion is not intended to be construed as an endorsement by either GSC or EPA for any such product.

HOW TO USE THIS MANUAL

This is a NEW User's Guide for the Graphical Exposure Modeling System (GEMS). New enhancements to GEMS through the spring of 1988 are reflected in this document, which is intended to replace all previous versions.

Chapter One gives you a broad introduction to GEMS and how it works. You'll learn how to communicate with the program, and what to expect for your efforts. You will want to read this section since it contains previously un-presented information. After reading this section you should be ready to use the system. If you are not experienced in GEMS use, you may want to read the additional chapter which describes the GEMS procedures you want to use.

A number of useful tables will be provided at the end of the completed volume, including information which is refered to in a number of places in the text. These tables provide information which is useful to a GEMS user but which did not fit in at any particular place in the user's guide. This guide includes a topical index to help you quickly find just the information you seek.

Many people have contributed greatly to the compilation of your user's guide. We would like to thank all of them for their efforts, especially the following people, (listed alphabetically) without whose individual efforts the project would not have been completed. Andy Battin, Jon Chen, Nancy Chiu, Terri Grasso, Loren Hall, Pat Harrigan, Dr. Russ Kinnerson, James Pillotte, Scott Rheingrover, Kellie Scialabba, Cathy Turner, and Stuart Wollman.

If you have comments on the manual, please drop us a line and share them with us. You can use the User Response Form at the end of the Guide or drop us a line directly. Its your feedback that will direct the publication of the next GEMS User's Guide.

Herb Hunt
Dalroy Ward
GSC, May 1988

Foreward

The Graphical Exposure Modeling System (GEMS) was first developed in 1981. The first version of GEMS consisted of only a few modeling tools with limited application, and of potential use to only a few scientists. Today, GEMS has evolved into a system that combines the power of over 40 different modeling and statistical analysis tools into one easy-to-use system. GEMS is used by over 250 scientists working at EPA headquarters and its regional offices, other Federal agencies, and State and local environmental and human health departments.

Several proprietary systems have been incorporated into GEMS including the powerful analysis and graphic display tool, Statistical Analysis System (SAS). These separate systems each has a user's guide available for those users who are not familiar with their individual operation. In addition, each of the environmental models in GEMS has its own user's guide. In this document, we have provided only brief overviews of the function of each of these separate systems, along with a quick reference to commands utilized within each. The reference section of this guide lists the model and system specific user's guides.

This publication has evolved from several previous writings by many different authors. We hope to provide on these pages a clear and concise guide to GEMS use for the user who already has a basic understanding of the functioning of the various component systems. *Editorial comments are not only welcomed, but encouraged.*

Every attempt has been made to insure that the contents of this users guide are correct and presented in the most useful manner. However, GEMS is an ever evolving system, and the possibility exists that even while we were writing these pages, the system may have outgrown the paper. We apologize, and hope that the next update will reach you soon. If you find documentation bugs, take a moment to report them to us by using the GEMS trouble reporting procedures outlined in section 8.1. Thanks!

This page intentionally blank.

Table of Contents

Chapter 1 - Introduction 1-1

1.1 Who Uses GEMS 1-1

1.2 How GEMS Works 1-1

1.3 GEMS on the EPA VAX Cluster 1-4
- 1.3.1 Getting to the Port Selector 1-4
  - 1.3.1.1 Washington DC - Direct Access 1-5
  - 1.3.1.2 Regional Dial Access 1-5
  - 1.3.1.3 DECnet ACCESS 1-6
  - 1.3.1.4 TYMNET 1-7
- 1.3.2 The RTP Port Selector 1-9
- 1.3.3 Running GEMS 1-14
- 1.3.4 Things to Know 1-14

1.4 Using GEMS 1-16
- 1.4.1 GEMS Menus 1-16
  - 1.4.1.1 Navigational Menus 1-16
  - 1.4.1.2 Parameter Entry Menus 1-18
- 1.4.2 Entering GEMS 1-21
- 1.4.3 GEMS Datasets 1-25
- 1.4.4 Exiting GEMS 1-29
- 1.4.5 TGEMS 1-30
- 1.4.6 Reporting Problems 1-30

1.5 Commands Summary 1-30

1.6 Communication with PCs 1-33
- 1.6.1 PCGEMS 1-33
- 1.6.2 Communications methods to GEMS 1-35
  - 1.6.2.1 Communications with Kermit 1-35
  - 1.6.2.2 Communications with CrossTalk 1-37
  - 1.6.2.3 Communications with PCPLOT 1-39

This page intentionally blank.

Table of Contents (Cont'd)

Chapter 2 - Chemical Property Estimation **2-1**

2.1 Describing Chemical Structures 2-1
2.1.1 SMILES Notation 2-2
2.1.2 Molecular Structure Entry (SFILES) 2-5

2.2 Retrieving Stored SMILES Notations (SMIGET) 2-5
2.2.1 SMIGET Input/Output Specification 2-6
2.2.2 Input dataset designation 2-8
2.2.3 Retrieve SMILES notation 2-9

2.3 Draw SMILES Structure 2-10

2.4 Octanol Water Partition Coefficient (CLOGP3) 2-11
2.4.1 The UDRIVE Access System 2-12
2.4.2 Using UDRIVE 2-13
2.4.3 Introd. to Octanol Water Partition Coefficient 2-14
2.4.4 Using CLOGP3 2-15

2.5 Chemical Property Estimation (CHEMEST) 2-16
2.5.1 Using CHEMEST 2-17
2.5.2 Estimating Property Values 2-21

2.6 New Chemical Property Estimation (NEWCHEM) 2-30

2.7 Automatic Chemical Property Estimation (AUTOCHEM) 2-31
2.7.1 AUTOCHEM Procedures 2-31

2.8 Fate of Atmospheric Pollutants (FAP) 2-38
2.9.2 Running FAP 2-42

Chapter 3 - Environmental Modeling **3-1**

3.1 Air Models (AIR) 3-2
3.1.1 GAMS 3-3
3.1.1.1 ISCLT 3-6

Table of Contents (Cont'd)

3.1.1.2 TOXBOX 3-7
3.1.1.3 Input 3-7
3.1.1.4 Output 3-8
3.1.2 INPUFF 3-11
3.1.2.1 Input 3-11
3.1.2.2 Output 3-13
3.1.3 PTPLU 3-13
3.1.3.1 Input 3-14
3.1.3.2 Output 3-15
3.1.4 BOXMOD 3-16
3.1.4.1 Input 3-17
3.1.4.2 Output 3-17

3.2 Soil and Groundwater Models (SOIL) 3-17
3.2.1 Soil - SESOIL 3-18
3.2.1.1 Input 3-19
3.2.1.2 Output 3-20
3.2.2 Soil - PRZM 3-22
3.2.2.1 Input 3-23
3.2.2.2 Output 3-23
3.2.3 Groundwater - AT123D 3-23
3.2.3.1 Input 3-26
3.2.3.2 Output 3-26
3.2.4 Groundwater - SWIP 3-28
3.2.4.1 Input 3-28
3.2.4.2 Output 3-29

3.3 Surface Water Models (WATER) 3-31
3.3.1 Surface Water - EXAMS II 3-31
3.3.1.1 Input 3-32
3.3.1.2 Output 3-33

3.4 Multi-Media Models (MULTI) 3-35
3.4.1 Environmental Partitioning Model (ENPART) 3-35
3.4.1.1 Input 3-36
3.4.1.2 Output 3-37
3.4.2 TOX-SCREEN 3-37
3.4.2.1 Input 3-38
3.4.2.2 Output 3-39
3.4.3 MICROBE-SCREEN 3-40

Table of Contents (Cont'd)

3.4.3.1 Input 3-40
3.4.3.2 Output 3-41
3.4.4 UTM-TOX 3-41
3.4.4.1 Input 3-43
3.4.4.2 Output 3-43

Chapter 4 - Geodata Handling 4-1

4.1 Geographic Data Lookup and Display (GEOLIST) 4-2
4.1.1 Site Level Retrieval of Data (SITERET) 4-4
4.1.2 Access Census Data (CENSUS) 4-10
4.1.3 Determine County Coverage (COVERAGE) 4-16
4.1.4 Geographic Data Management (GEODM) 4-19
4.1.5 HUCODE/SOIL locator (HUCODE) 4-23
4.1.6 Convert Geocode to Lat/Long (LATLON) 4-27
4.1.7 Lookup/Examine STAR Station Data (STAR) 4-31
4.1.8 Find US Cities (USCITY) 4-34
4.1.8.1 All Cities in a State 4-36
4.1.8.2 Selected County of a Selected State 4-37
4.1.8.3 Selected City in a Specified State 4-38
4.1.8.4 All Cities Within a Specified Radius 4-40

4.2 Geodata Handling Mapping Procedures 4-41
4.2.1 Draw Block Group/Enum. Dist. Map (BGEDMAP) 4-42
4.2.2 Draw County Map (CNTYMAP) 4-44
4.2.3 Hydrological Unit Map (HUMAP) 4-56
4.2.4 State/County Map Procedure (STCOMAP) 4-62
4.2.4.1 County Map 4-63
4.2.4.2 State Map 4-70
4.2.4.3 Mapping Options 4-75

Chapter 5 - GEMS File Management Facility 5-1

5.1 Create Datasets (CREATE) 5-3
5.1.1 Create Datasets by Data Entry 5-3

Table of Contents (Cont'd)

5.1.2 Use VT100 Dataset Editor (VTCREATE) 5-6
5.1.3 Sort a Dataset (SORT) 5-9
5.1.4 Merge Two Datasets (MERGE) 5-10
5.1.5 Concatenate Two Datasets (CONCAT) 5-12
5.1.6 Data Transformation (TRANSFORM) 5-14
5.1.7 Extract Data From a Dataset (EXTRACT) 5-15

5.2 Modify Datasets (MODIFY) 5-18
5.2.1 Edit a Dataset (EDIT) 5-18
5.2.2 Use the VT100 data editor (VTMODIFY) 5-19
5.2.3 Add Records (ADDREC) 5-20
5.2.4 Delete Records (DELREC) 5-21
5.2.5 Modify Dataset Attributes (ATTRIBUTE) 5-22

5.3 Delete Datasets/Graphs/Tables/Maps/Reports (DELETE) 5-23

5.4 Output Datasets/Graphs/Tables/Maps/Rep. (OUTPUT) 5-24
5.4.1 Review a Dataset (DATASET) 5-25
5.4.2 Review a Graph (GRAPH) 5-27

5.5 Export Datasets/Graphs/Tables/Maps/Reports (EXPORT) 5-27
5.5.1 Convert a GEMS Dataset to a VAX Format (G2VAX) 5-28
5.5.2 Convert a Graph to a Graphics File(G2GRAPH) 5-30

5.6 Import Datasets/Graphs/Tables/Maps/Reports (IMPORT) 5-31
5.6.1 Convert a VAX File to a GEMS Dataset (VAX2G) 5-32
5.6.2 Convert a SAS File to a GEMS Dataset (SAS2G) 5-33
5.6.3 Con. a Graphics File to a GEMS Dataset(GRAPH2G) 5-34

5.7 Change Status of Datasets/Graphs/Tables/Maps/Reports (STATUS) 5-35
5.7.1 Save Temporary Dataset (SAVEDATASET) 5-36

Chapter 6 - Statistics **6-1**

6.1 Contingency Tables (CT) 6-1

6.2 Descriptive Statistics (DS) 6-3

Table of Contents (Cont'd)

6.2.1 Simple Statistics (MEAN) 6-3
6.2.2 Univariate Statistics (UNIVARIATE) 6-4
6.2.3 Percentile Statistics (PERCENTILE) 6-5

6.3 Simple Regression (SR) 6-7

6.4 Multiple Regression (MR) 6-8

Chapter 7 - Graphics 7-1

7.1 Barchart (BAR) 7-2

7.2 Histogram (HIST) 7-9

7.3 Scattergram (SCAT) 7-14
7.3.1 Single Plot 7-16
7.3.2 Plot2 7-19
7.3.3 LINREG 7-21
7.3.4 Overlay 7-23

7.4 Isopleth (ISO) 7-25

7.5 Rose (ROSE) 7-31
7.5.1 Produce Rose Using Lat/Long Coordinates 7-33
7.5.2 Produce ROSE with Manual Input (ROSEMAN) 7-35

Chapter 8 - Utilities 8-1

8.1 Report GEMS Problems (PR) 8-1

8.2 Invoking the Statistical Analysis System (SAS) 8-2

8.3 Convert ASCII to SAS Files (ASC2SAS) 8-3

Table of Contents (Cont'd)

8.4 Invoke ARCINFO (ARCINFO) 8-4

8.5 Check CPU Usage (TOPCPU) 8-4

8.6 Directory Management Utilities (DIRMAN) 8-4
8.6.1 Get SFILES Structure Files (GETSFILES) 8-5
8.6.2 Delete SFILES Structure Files (DELSFILES) 8-5

8.7 VAX/VMS Utilities (VMSUT) 8-5
8.7.1 Show Disk Quota (QUOTA) 8-6
8.7.2 List Directory 8-6
8.7.3 Edit a VMS File (VMSEDIT) 8-7
8.7.4 Access VMS Mail (VMSMAIL) 8-7
8.7.5 Purge VMS files (VMSPURGE) 8-7
8.7.6 Delete VMS Files (VMSDELETE) 8-8

Chapter 9 - Information and News 9-1

9.1 Send Messages to GEMS Staff (FEEDBACK) 9-1

9.2 Type GEMS News Message (NEWS) 9-2

9.3 GEMS Bulletins (BULLETINS) 9-3

Chapter 10 - Datasets 10-1

10.1 GEMS Datasets 10-1
10.1.1 Datasets to Support Environmental Models 10-2
10.1.1.1 Population Density and Exposure Estimates 10-2
10.1.1.2 Atmospheric Assessments 10-4
10.1.1.3 Soil and Groundwater Assessments 10-5
10.1.1.4 Surface Water Assessments 10-11
10.1.1.5 Multi-Media Assessments 10-13
10.1.2 Datasets Identifying Chemical Release Sites 10-16
10.1.3 Datasets with Chemical Information 10-23
10.1.4 Geographic and Cartographic Datasets 10-24

Table of Contents (Cont'd)

810.2 Auxiliary Files 10-28
10.2.1 Datasets to Support Environmental Models 10-28

Tables **T-1**

1 States (Names, USPS Codes, FIPS codes T-2

2 Phone Numbers T-4

References **R-1**

Index **I-1**

Chapter 1 - Introduction

The amount of data on which policy decisions are based has created an ever increasing burden for planners. There needs to be a way to evaluate a large volume of information quickly and accurately. To answer this need, the Office of Toxic Substances (OTS) has developed the Graphical Exposure Modeling System (GEMS). GEMS is an interactive information management tool designed to allow the quick and meaningful analysis of environmental problems. Figure 1-1 illustrates the functional relationship of the various GEMS components.

1.1 Who Uses GEMS

GEMS was originally designed for use by EPA scientists and modelers who needed a tool to provide comprehensive policy planning support to EPA administrators. It has subsequently evolved into a tool for use by a wider range of environmental professionals, including those at the state level, as well as other federal agencies such as the Department of Energy. Plans are now in the works to increase the scope of GEMS use by making it available to local communities and several international agencies. The environmental scientist at virtually every level from federal to local community will eventually be able to use this tool to provide policy and planning support for a broad range of environmental issues.

1.2 How GEMS Works

Like most scientists elsewhere, the EPA turned to the computer to help solve their number crunching tasks. GEMS was developed to provide interactive analysis of information in a

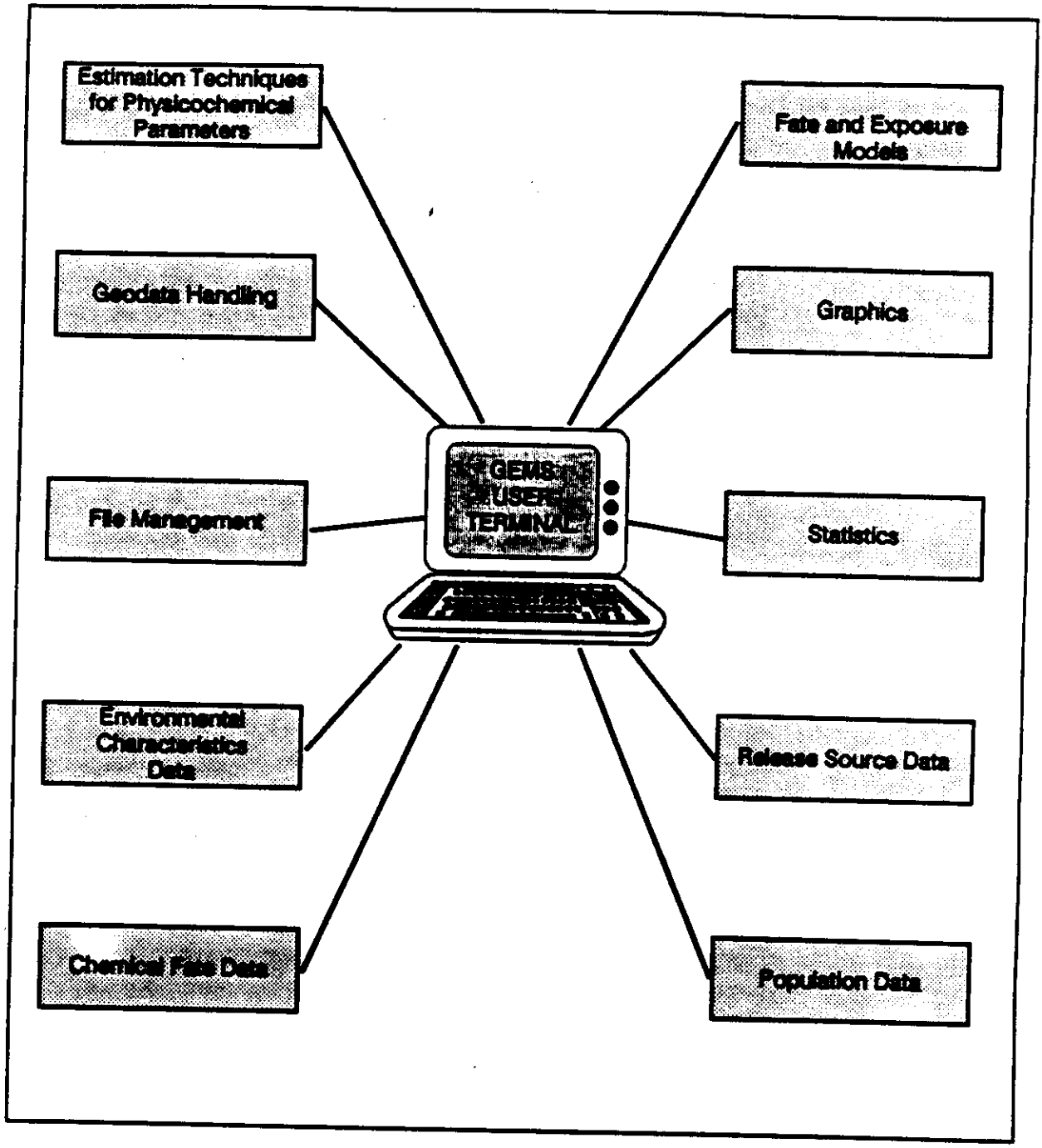

FIGURE 1-1. GEMS Components

broad range of databases of interest to scientists and policy makers. Data such as census information on population distributions are used in studies with other data such as chemical fate or modeling data. Often many databases are utilized in producing GEMS outputs.

GEMS is an operating environment that ties together several previously discrete analysis tools into a coordinated system that allows for multiple types of analysis. These discrete systems are handled under a common environment that makes a user's data available to multiple individual modules of GEMS. Data produced by certain GEMS models, for example, could then be utilized to provide graphic analysis.

GEMS allows the user to:

- Estimate chemical properties.
- Assess the fate of chemicals in receiving environments.
- Model resulting chemical concentrations.
- Determine the number of people potentially exposed.
- Estimate the resultant human exposure and risk.

While some modeling knowledge is required to operate the models, the use of GEMS as the host environment for accessing these models is extremely simple and requires only minimal computer skill.

The modular design of GEMS supports the implementation of additional features on an ad hoc basis. Modules already implemented include:

- Molecular Structure Entry
- Physicochemical Property Estimation
- Fate of Atmospheric Pollutants
- Air Models
- Soil Models
- Water Models
- State/County Map
- Geographical Data Manipulation and Mapping
- Census Data Retrieval

HELP is always available to the user at the touch of a key, and even an AUTOHELP facility is available to assist new users,

which provides help messages every step of the way. More experienced users may elect not to use the menu navigation system by simply issuing direct commands to GEMS to access specific functions. Menus are used to set up models and provide information input to GEMS. There is also a "prompt mode" available for users who feel more comfortable in an interactive dialog rather than menu selection. This range of user options provides broad support to users of any skill level.

1.3 GEMS on the EPA VAX Cluster

GEMS resides on the VAX Cluster of computers maintained by EPA at Research Triangle Park (RTP), North Carolina. The VAX Cluster currently consists of two VAX 8600's and one VAX 11/785. You must have an account on this system to access GEMS. If you do not currently have an account contact:

Ms. Cathy Turner at (202) 382-3929 / FTS 382-3929

She will send you information on opening an account. When you receive your account confirmation from the National Computer Center (NCC) at RTP be sure to keep it in a safe place. When you log onto the system, you will need to know the Username you were assigned, the initial Password for your Username and the Project to which your work will be charged.

1.3.1 Getting to the Port Selector

With a valid account on the RTP system you may use your modem and terminal or PC to log onto the system. The actual connection to the VAX Cluster is accomplished by using one of the following access methods:

1. Washington, DC Direct Access
2. Regional Dial Access
3. DECnet Access
4. TYMNET Access

You will need to investigate which of these alternatives is best suited for your needs. Some incur additional charges, while some are toll-free. These methods are discussed in the following sections. If you are in metro Washington, D.C., or in

the area of one of the regional access numbers, TYMNET access is not available.

Yet another access method is available for international users. More about international access is discussed in section 1.3.1.4.

1.3.1.1 Washington DC - Direct Access

From the metropolitan Washington DC area, you access the Cluster by dialing the local access number for the communications line to RTP, 488-3671. This number will allow you to access the port selector at RTP. TYMNET nodes in the metropolitan DC area can not access the RTP computer, you must use one of the direct access numbers. 488-3671 supports both 1200 and 2400 baud,or 488-1515 for 1200 baud only. For calls from DC, your communication parameters should be 7 data bits, even parity. For information on what happens after you dial this access number, go to section 1.3.2 for instructions on use of the port selector.

1.3.1.2 Regional Dial Access

To use the Regional Data Switch, dial the appropriate number listed in Table 1-1 below.

These numbers should be used if you are working in an EPA regional office or either of the two EPA labs, or if these numbers are local calls to you. When calling some of these regional data switches, you may be prompted for a device type. For example, the Region II Network users must first select for ASCII 1200 baud. Generally, if you have an IBM PC or a VT100 emulator, respond to terminal-typep rompts with An alternative is to simply hold down the control key and press the key five times. You will get a message welcoming you to the system and asking for a system choice. These choices will include PDP, NCC, and TCP.

To access the VAX Cluster, the home of the GEMS system, select This will provide you with access to the NCC IBM through the protocol converter. Regions V and VI will be

prompted with **"Commtex Data Exchange. Enter terminal's device type:"**. Type [redacted] for help and respond to the following prompts according to the terminal type you are using. The next message you will see will ask you to select the mainframe you want to use, select [redacted] After this point, the process will be identical to the procedures described in section 1.3.2, The RTP Port Selector. For more detailed information on what to do next, go to this section.

TABLE 1-1. Regional Dial Access Numbers for 300 and 1200 Baud Operation as of 3/1/88

REGION	OFFICE LOCATION	TEL. NUMBER	FTS
I	Boston, MA	(617) 565-3774	8 835-3774
II	New York, NY	(212) 264-1554	8 264-1554
III	Philadelphia, PA	(215) 597-0103	8 597-0103
IV	Atlanta, GA	(404) 347-7285	8 257-7285
V	Chicago, IL	(312) 353-2468	8 353-2468
Lab	Cincinnati, OH	(513) 569-7821	8 684-7821
VI	Dallas, TX	(214) 655-6409	8 255-6409
VII	Kansas City, KS	(913) 371-8118	8 757-8118
VIII	Denver, CO	(303) 293-1400	8 564-1400
IX	San Francisco, CA	(415) 495-0168	8 454-0168
Lab	Las Vegas, NV	(702) 798-3101	8 545-3101
X	Seattle, WA	(206) 442-4423	8 399-4423

1.3.1.3 DECnet ACCESS

If you have a local VAX which is on the EPA DECnet network, you can access the VAX Cluster across the network. To do so, log into your local VAX and then use the VAX/VMS SET HOST command to connect to the Cluster.

[redacted] for access via DECnet

or

[redacted] for access over the X.25 network

Your system must have DEC PSI available, and your system manager must have set up the address file in order to use the

X.25 link. Check with your system manager to determine if DEC PSI is available and set up on your system.

If your network link is at 9600 baud or less you should use the X.25 network for access to the Cluster. At such low speeds, the DECnet overhead can make editing a tedious process at best. By using this method, you bypass the port selector and connect directly to the Cluster.

1.3.1.4 TYMNET

TYMNET is available if you are outside of the cities with regional offices listed in Table 1-1 and without access to toll-free numbers. To use TYMNET, you will need to call NCC at (919) 541-4506 (FTS 629-4506) to find the number for your area. When you call NCC to get the phone number, make sure you ask for the number that supports the speed at which you plan to communicate (1200 or 2400 baud). When you are using TYMNET, your terminal should be set to 8 data bits, no parity. When you dial the TYMNET number, TYMNET will respond with:

- **Enter your terminal id:** (at 300 baud)
- **$xx%!@xxxx$&x%x%#** (at 1200 baud)
  (The garbage may be a couple of lines.)
- You will see nothing immediately following carrier detection at 2400 baud.

If you are calling at 1200 baud, don't worry about the garbage, that's normal. In all cases, enter an *A* in response. You do not have to wait until the text is finished before typing [redacted] At 2400 baud, you will see nothing until you enter [redacted]

You will then see:

Please Log In:

Respond with:

[redacted] for 1200 baud

or

EPA2400 for 2400 baud operation

If you are using a 2400 baud line, you will need to know the TYMNET password. To learn the TYMNET password for 2400 baud operation, you should contact NCC at (919) 541-4506.

Once you have successfully gotten to this point, the TYMNET network will advise you of a successful connection with a message like:

NODE xxxx HOST 1038 EPA IS ONLINE

Congratulations, you are now connected to the port selector at RTP. Continue with the procedure discussed in section 1.3.2 below.

If you are having trouble logging into the system, and believe you are doing everything right, there may be a problem with the computer itself. The National Computer Center maintains a status phone to let users know if there are problems with the system. You can call (FTS) 629-2969, or (919) 541-2969 to check on the current system status.

TYMNET International Access

To obtain information for accessing the NCC computer from overseas, first dial your local TYMNET node. At the **Please Log In** prompt, type [redacted] to enter the TYMNET Information Service. At the Main Menu select the entry for International Access.

TYMNET provides access to and from over 70 foreign countries, and you can obtain a listing by selecting #1 from the second menu. For detailed information on a specific country that is on the list, select option #2. You will be prompted to enter the name of the country, and then will be given information including a contact in that country (name, address, telephone number, and telex) and technical information including access, speeds and protocols supported, and prices. Inbound and outbound rates are listed

in Menu Item #4; customer support information is provided in option #7.

TYMNET Customer Support and Trouble-Shooting services are available to help in placing an international call. The toll-free number is (800) 336-0149. When trying to locate NCC's international address you may be asked to contact a communications consultant in external network services at (703) 356-6993.

1.3.2 The RTP Port Selector

Once you have connected to the port selector, you are almost there. Enter one or two carriage returns to bring up the port selector menu. (If you enter too many returns, you will be disconnected). The menu will look like this:

```
WELCOME TO THE ENVIRONMENTAL PROTECTION AGENCY
HEADQUARTERS DATA SWITCH

TO ACCESS:                          TYPE:
IBM (TTY)                           NCC, IBMPSI
IBM 3270 EMULATION                  NCC,TCP
EPA/DIALCOM ELECTRONIC MAIL         EMAIL
PRIME                               system name
VAX                                 NCC,system name
MODEM POOL                          MODEM,999-9999
OTHER SERVICES                      HELP
                NEED HELP? TYPE HELP
YOUR SELECTION >
```

FIGURE 1-2. Initial Messages

You have two choices of VAX systems at the prompt:VAXA and VAXB. You may enter [redacted] to call the node named VAXTM1 which gives you access to SAS and FOCUS software. If you plan to use any of the graphical capabilities (all of which use SAS) in GEMS, you should select VAXA. VAXB calls the node named CASTOR which gives you access to ARCINFO software. Enter [redacted] to call CASTOR. Once you tell the port selector to connect you to the VAX, you will see:

Connected.

Now enter a few carriage returns (one or two) until you get the computer's attention. It will then respond with:

Username:

Enter the username which was assigned to you by the RTP accounts personnel. You will then see:

Password:

Enter the password for your username. If this is your first time logging in, then enter the password which was given to you on your account authorization information from RTP. Otherwise enter your current password. If this is your first time logging on, remember to change your password when you finish logging in.

After entering your password, if you see **User authorization failure**, then you have entered either your username or password incorrectly. The system will re-prompt you for your username and password. If you enter everything as entered on the account authorization information you received from RTP, then give User Support a call (800) 334-2400, or (FTS) 629-7862 for assistance in solving the access problem.

```
*****************************************************************
VAX USER SUPPORT: (FTS)629-7862 or 919-541-7862 or 800-334-2405
*****************************************************************
VAX 8600 Operations Status Phone: (FTS)629-2969 or 919-541-2969
For the Current Operations Schedule Type:  OPERATIONS_SCHEDULE
*****************************************************************

Last interactive login on ...
Last non-interactive login on ...
*****************************************************************
                    CURRENT NEWS ALERTS
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

************TYPE "NEWS ALERT#" TO VIEW AN ALERT**************
```

FIGURE 1-3. VAX Cluster Messages

Project:

At this point you enter the project which has been assigned to you, it will be listed under "account" on your account authorization sheet. You cannot finish the login procedure without this information. If you cannot remember your project then type [redacted] to log off of the system and find the account authorization information that you received from RTP to confirm that your account had been set up. This will have the project information.

Once you have completed the process of logging into the system you will see the VAX/VMS prompt, the dollar sign ($). This indicates that the system is ready and waiting for input from you. Under some circumstances you may have a system prompt of the node ID in the network (i.e.: CASTOR> or VAXTM1 >) rather than the "$." If your account is setup in this manner, then substitute your prompt for the "$" we use in this manual for illustration.

VAX/VMS uses the Digital Command Language (DCL) as the interface between you and the VAX. One of the most useful of the DCL commands is the "HELP" command. This will access the VAX/VMS on-line help function. You can simply type [redacted] or type [redacted]

For example you can enter [redacted] for help on the copy command. You can get help on any DCL command in this manner. To give you a hand, if this is your first time on a VAX, we present several useful concepts below.

If this is your first time logging into your account, you should IMMEDIATELY change your password to something you won't forget. The VAX requires this change for security reasons. If you forget to change your password, your username will be locked against use once you log off after your initial session. Then you will have to call USER SUPPORT to have your Username unlocked. When changing your password try to choose something which only you can guess. Avoid using words that others might guess such as your spouse's or child's name, your middle name, or other words that are easily associated with you. Remember you are responsible for

guarding the security of your password. Memorize the password and do not keep it written down next to your terminal or in your office. To help insure the security of your password, the system will not display your password as you enter it. Looking over your shoulder will not allow others to see your entry.

To change your password, enter:

```
$ SET PASSWORD
```

You will be prompted for your old password, new password, and verification: once you have entered your currently valid password, you will be asked to enter the new password that you want to use in the future. To verify your entry of the new password, you will be asked to enter it a second time. BE SURE TO REMEMBER YOUR PASSWORD. IF YOU FORGET IT, YOU WILL LOSE ACCESS TO YOUR ACCOUNT.

When you log in to the system you will be in your main directory. At this point you can edit files, create sub-directories, or run a program, for example: GEMS. Remember, however, that if this is your first time logging on, you must set up the GEMS environment before you can run GEMS. Be sure to learn about creating sub-directories as they are very useful for organizing your files. If you put everything you do into your main directory, it will get very cluttered and it will be difficult to find the files you are interested in. In the next two paragraphs we'll introduce a few DCL commands to you and give a short description of how file names work on the VAX. If you are new to VAX computers, please be sure to read both of these sections.

File names under VAX/VMS have two components, the file name and the file type in the format:

filename.filetype

The file name can have from one to 39 characters chosen from the letters A through Z, the numbers 0 through 9, and the underscore (_) and dollar sign ($). The file type can be from 0 to 39 characters from the same set of characters. The file

name and the file type are separated by a period (.). Some examples of legal file names are:

LOGIN.COM	DATA.
GENERAL.DAT	GAMS_ISC01.CONC

The system recognizes a number of default file types for special purposes. Use of these file types should be limited to the types of files they are meant for:

FILE TYPE	USE
COM	DCL Command procedure (example: LOGIN.COM)
DIR	Directory or subdirectory
EXE	Executable file
OBJ	Object file created by a compiler or assembler
FOR	Source file for FORTRAN compiler
MAI	VAX/VMS Mail file

In addition to a file name and type, every file has a version number assigned to it when it is created or revised. When you initially create a file it has a version number of 1. As you modify the file the version number is increased by one each time. The system will maintain all versions of a file. You should use the PURGE command to delete all versions but the highest numbered (most current) one.

A wildcard character is a symbol that you can use with many DCL commands to apply the command to several files at once, rather than specifying each file individually, or to select a sub-set of the available files in a directory. The wildcard characters are the asterisk (*) and the percent sign (%). The asterisk matches any number of characters in a file name or file type:

```
DIRECTORY *.DAT
```

DIRECTORY *.DAT would give you a list of all files in your directory with the file type of DAT. The percent sign matches

only a single character in the position at which it is specified:

```
DIRECTORY CHAPTX.TXT
```

DIRECTORY CHAPT%.TXT would list out files such as CHAPT1.TXT, CHAPT2.TXT, but would not list CHAPT.TXT or CHAPTIX.DAT.

1.3.3 Running GEMS

All you will need to do to run GEMS is enter:

```
$ GEMS
```

See section 1.4, Using GEMS, for additional information on using the system.

1.3.4 Things to Know

The first time you run GEMS, a number of files and directories will be created in your account. The most important of these is the sub-directory GEMSLIB. This directory will be created by GEMS under your primary directory. If your primary directory was [GEMSUSER] then you would have the new directory [GEMSUSER.GEMSLIB]. This directory is very important. It contains your GEMS user dataset library. You should never delete the files in this directory or the directory itself except from within GEMS. If you delete files within this directory, you lose access to that dataset and run the risk of damaging your user dataset library. If it is damaged, you will lose access to the rest of your data.

If you run the Estimation Procedure routines, you will have a number of files ending in the extension .SFS created in your main directory. These files are important to the execution of these programs and are used to store the structures which you are entering. If you delete these files, you will lose access to the structures which you have previously entered. If you no longer need access to previously entered structures, you can delete these files to conserve disk space.

Files will also be created, in your directories, as you run some of the models accessible through GEMS or when you use the

EXPORT feature under GEMS File Management. Files existing outside of the GEMS file management structure must be managed by you. You may save or delete these files as you deem appropriate.

Remember, if you have not accessed a file for 45 days, it will automatically be archived to conserve disk space, and will be available only after recalling it from archive. To learn how to retrieve a file from archive type [illegible] or consult the NCC VAX Users Guide.

Changing file protections can be very helpful in preventing the accidental erasure of your files. To change file protection, use the VAX/VMS SET PROT command:

SET PROT=(S=RWD,O=RWD,G=RWD,W=RWD)

FILENAME.EXT

You can change the protection for various categories of users:

S(ystem)	System level Usernames
O(wner)	Yourself
G(roup)	Usernames under the same project as yours
W(orld)	All other Usernames

Remember that you must remove a protection from all categories after yours. For instance, if you have removed "Delete permission from O(wner)" but you have left it in for "W(orld)," you will still be allowed to delete the file.

The protection codes are:

E to be allowed to execute a file. Mainly applies to .COM and .EXE files. Allows a user to execute the file without being able to read or write the file.

R for read permission. Users in the applicable category will be allowed to read the file. This level of protection

also allows a user to make a copy of the file.

W for write permission. Users will be allowed to write new information to the file. Be very careful in allowing this level of protection.

D A user will be allowed to delete the file. Omit this protection to prevent yourself or others from being able to inadvertently delete this file.

1.4 Using GEMS

If you have successfully completed section 1.3 of this manual, you should now have logged on to the computer system and have access to GEMS. Before you actually run and use GEMS, there are some concepts you should keep in mind.

1.4.1 GEMS Menus

GEMS comes with two basic types of menus, navigational menus and parameter editing menus. Knowing the differences between the two will be helpful in understanding how the system works and how to use the system most efficiently.

1.4.1.1 Navigational Menus

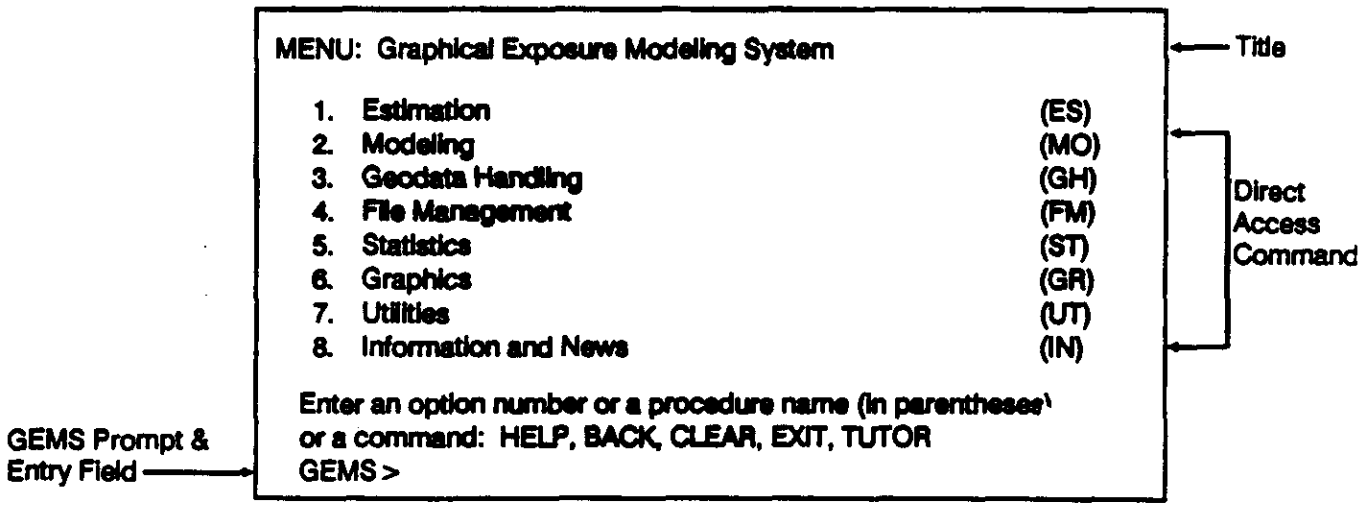

FIGURE 1-4. Typical Navigational Menu

Navigational Menus help you navigate your way through GEMS. They present you with a list of options from which to select. When you select one, you will proceed to either another navigational menu or a parameter editing menu. To get to many GEMS functions, you may have to go through several navigational menus.

Title:	This is the title of the menu. If you have problems at a menu, writing down the procedure you were using (the title of the menu) can greatly assist the support staff in solving any problem.
Reference number:	Each parameter on the menu will have an associated reference number. Enter this number to select the operation or procedure.
Description:	The name of the operation that is associated with the number preceding it is presented.
DAC:	The Direct Access Command for the procedure or operation. (See below for a discussion of DAC's).
Entry Field:	This is the area of the screen where you will enter your function selection.

To select an option from the menu, you need only type the reference number associated with that item.

Once you are familiar with GEMS, you can issue a Direct Access Command (DAC) for the functions that you want to use. The Direct Access Command lets you name a destination without relying on the system navigational menus for guidance. This can save you time and keystrokes when you know exactly what you want to do and the proper command to do it. Remember, however, that DAC's are available only from navigational menus.

At navigational menus, you can type HELP to bring up a help screen that explains your current menu. You can also type HELP n where n is the reference number, and you will get specific help about that selection.

The DAC feature is very powerful and useful, but you should take care in its use. The problem with relying on the Direct Access Commands too heavily is that you miss many of the other useful features which are available through GEMS. After navigating through the menu screens, you will reach the procedure you desire. Once there, you will use parameter entry menus.

1.4.1.2 Parameter Entry Menus

Parameter Entry Menus are the menus which are used to enter values into GEMS for specific parameters needed by a GEMS procedure in its execution. Several conventions exist in entry menus which you should be aware of. A field which contains an asterisk (*) will have a default value applied if you do not change it. Many menus used in setting up input for the models used in GEMS have default values already in place in the parameter fields. These values should only be used if you have verified that they make sense for your application. While an attempt has been made to select values which will apply in many situations, it cannot be assumed that the default value will apply in any specific situation. Finally remember that DACs will not work at a parameter entry menu.

Title → MENU: Monthly Climate Data for Year 1

ref	parmname	parameter description	Value	Iindex
1.	TA	Air Temperature (deg C)	15.08	(1)
2.	NN	Cloud Cover (frac.)	0.77	(1)
3.	S	Relative Humidity (frac.)	0.84	(1)
4.	A	Short Wave Albedo (frac.)	0.19	(1)
5.	REP	Evapotranspiration (cm/day)	0.00	(1)
6.	MPM	Rain Depth (cm/month)	13.86	(1)
7.	MTR	Mean Storm Duration (days)	0.24	(1)
8.	MN	Number of Storm Events (#/mnth)	11.00	(1)

Reference Numbers

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT

GEMS Prompt & Entry Field → GEMS >

FIGURE 1-5. Typical Parameter Entry Menu

Title:	This is the title of the menu. If you have problems in GEMS, writing down the procedure you were in and the title of the menu you were at can greatly help in finding and solving the problem.
Reference Number:	Each parameter on the menu will have a reference number associated with it. This is the number which you will use to specify which parameter you want to assign a new value to.
Parameter Name:	The name of the parameter. Every attempt has been made to associate a meaningful name with the parameters.
Description:	A short description of the parameter.
Value:	This field displays the current value of the parameter. Many parameters will be initially displayed with no value. You must specify a value for all items on the menu which do not have default values.
Index:	This lets you know if a parameter is a single value or an array parameter. If this field is blank, it is a single value. If the parameter is an array parameter (i.e., it can take on more than one value), this field will initially have a "(1)" displayed there. This indicates that the value of the first index is being displayed on the menu.
Entry Field:	This is the area of the screen where you will enter your information to fill the editing screen.

Now that you know a bit about the menu, here's how you modify parameter values in the menu.

To modify a parameter you type its reference number followed by a space and then the new value: 1 NEW VALUE If you are using a DEC video terminal or Tektronix 41xx series terminal, the new value for the parameter will be displayed on the menu, after a CR, to indicate that the value has been changed. If you are entering values for an array parameter then you should enter just the reference number. You will then be prompted for the values for each element of the array. You end the process by entering just a carriage return for a value. You will then be returned to the same menu and the value of the last element of the array will be displayed and the number of elements in the array will be indicated by the index field.

You can use the SHOW command to examine the value of a specific element of the array. Use is simple, just type the command SHOW PAR-NAME INDEX where you replace PAR-NAME with the name of the parameter, and INDEX with the number of the element you want to see. The value and index fields will be updated with the information you requested.

By typing HELP you can bring up a general help screen about the menu, or by typing HELP n where 'n' is the reference number, you can bring up help for a specific parameter on the menu.

When entering data you may also put more than one reference number on a line. Separate the entries using commas (*1 new value,2 another new value,3 the third new value*) to modify the values for Items 1, 2 and 3.

Once you have finished entering values into this menu, you type NEXT to proceed to the next menu. If the menu you were on is the last in the sequence, then the NEXT command will proceed to run the program.

Should you decide that you want to end the procedure you are in prior to finishing your work, you can type END to end the current procedure and return to the previous navigational menu or CLEAR to end the current procedure and return to the main GEMS menu.

1.4.2 Entering GEMS

If you have set up your LOGIN.COM file as explained in section 1.3.2 then you will be able to simply enter the command *GEMS* to run GEMS. If you have not yet setup your LOGIN.COM file, turn to section 1.3.2 and do so before proceeding.

When you enter GEMS, it will print out the GEMS NEWS file to let you know about any new procedures or updates which have been made to GEMS recently. Once you have seen the GEMS NEWS, it's time to select your terminal type:

```
MENU: Terminal Type Specifications

   1.VT100-compatible terminal          2.Tektronix 4010 terminal
   3.VT100 with TEK41040 emulator       4.Tektronix 4014 terminal
   5.80 column ASCII terminal           6.Tektronix 4105 terminal
   7.132 column ASCII terminal          8.Tektronix 4106 terminal
   9.LA120 DECwriter terminal          10.Tektronix 4107 terminal

Please identify your terminal type by number
?
```

FIGURE 1-6. Terminal Selection Menu

Enter the number which corresponds to the terminal with which you logged onto the computer. This lets GEMS set up conversational parameters for such functions as full-screen display. This is important! Many GEMS features will only display on a particular terminal type (i.e., while you may run one of the mapping options on any terminal, a TEKTRONIX or compatible terminal is required to display the map). GEMS will still let you request these functions but will simply store the map or graphic which is created in your library for later viewing on the proper terminal. Once you have entered your terminal type GEMS gives you the official Welcome:

```
GRAPHICAL EXPOSURE MODELING SYSTEM

Version 8.1

Developed by
GENERAL SCIENCES CORPORATION

For
U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF PESTICIDES AND TOXIC SUBSTANCES
```

FIGURE 1-8. GEMS Welcome Screen

After a few seconds the main GEMS menu will appear.

```
MENU: Graphical Exposure Modeling System

  1. Estimation                                  (ES)
  2. Modeling                                    (MO)
  3. Geodata Handling                            (GH)
  4. File Management                             (FM)
  5. Statistics                                  (ST)
  6. Graphics                                    (GR)
  7. Utilities                                   (UT)
  8. Information and News                        (IN)
Enter an option number or a procedure name (in parentheses)
or a command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 1-7. GEMS Main Menu

You are now ready to begin work. From this menu you can either navigate through the available choices to the function that you want to use or use the menu DAC to jump there directly (as discussed above). The DAC is displayed at the far right of the selection screen. For instance, to advance to the Environmental Models section of GEMS you could either enter 2 from the main menu or enter [illegible]

Each of the items on the main menu is the subject of a chapter in this manual. Because of the complexities of Chemical Property Estimation and Modeling, separate User's Manuals have been written for these functions. In Chapters 2 and 3, we

will provide you with a broad overview of these operations and refer you to the appropriate detailed user's guide. In Chapters 4 through 8, we present a much more detailed discussion of GEMS procedures used in support of the modeling and estimation operations. For now, here's a short overview of each function.

Chemical Property Estimation

The Estimation option provides you with a variety of procedures which can be used to estimate a number of properties on a wide range of chemicals. These procedures fall into two classifications: Structure Entry (SFILES and SMILES programs) and Property Estimation (CLOGP, CHEMEST, NEWCHEM, AUTOCHEM, and FAP programs). More about Chemical Property Estimation can be found in Chapter 2.

Environmental Modeling

The GEMS Modeling functions provide access to a variety of computer models which simulate the effects of chemicals released to air, water, and soil. The models are grouped by media into the categories of AIR, WATER, SOIL, and MULTI-MEDIA. The AIR models deal with the fate of contaminants released into the atmosphere. The WATER models deal with the fate of a contaminant in surface water bodies, and the SOIL models address transport and fate of a contaminant in soil and ground water systems. The MULTI-MEDIA models deal with the transport and fate of a contaminant as it moves through the atmosphere, soil and surface water. All of these models produce estimated concentrations of the contaminant under study. Many of these models create GEMS datasets which may be used in subsequent analyses. More about Modeling can be found in Chapter 3.

Geographic Data

Geographic Data Handling allows access to mapping routines and to specialized data retrieval operations which access data by location (latitude/longitude) or geographic codes (such as

state name or State FIPS code, county name or County FIPS code, etc.). Please remember that the mapping operations are viewable only on Tektronix or compatible terminals at present. More about Geodata handling operations can be found in Chapter 4.

GEMS File Management

File Management allows for the creation and manipulation of GEMS datasets, some of which can be used as the input for other GEMS procedures. This capability also allows you to import data from outside of GEMS for use with GEMS procedures. More about the File Management Operation can be found in Chapter 5.

Statistics

The Statistics option provides the user with a variety of statistical analysis capabilities and functions. This capability has been implemented using routines from the International Mathematical and Statistical Library (IMSL) and from the Statistical Analysis System (SAS). More about Statistics can be found in Chapter 6.

Graphics

Selection of the Graphics option allows you to display data from the datasets in the system and/or your own dataset library. Many of the models installed in GEMS create GEMS datasets containing the results of the model's execution, which are usable by the GEMS graphics progams. This allows you to graphically represent information and modeling results to assist in analysis. Many of the graphics available through this operation have been implemented utilizing SAS Graphics capabilities. Most GEMS graphics require the use of a Tektronix compatible terminal, or a PC running Tektronix emulation software for viewing the graphs. More about Graphics can be found in Chapter 7.

Utilities

The Utilities operation allows access to a number of useful capabilities. Particularly important under this operation is the

Problem Reporting utility, allowing GEMS users to communicate with the GEMS system staff. This feature can be used to send problem reports, complements, and development suggestions and requests to the staff. A number of VAX/VMS utilities are also available under this GEMS Operation. More about Utilities can be found in Chapter 8.

Information and News

This operation allows you to convey messages to GEMS staff and for GEMS staff to inform you about new application or corrections or other newsworthy items. More about this operation can be found in Chapter 9.

1.4.3 GEMS Datasets

One of the most important concepts that must be mastered for a full understanding of GEMS is that of the GEMS Dataset. There are several types of GEMS Datasets. GEMS User Datasets are datasets which have been created by and are only accessable to a GEMS user. These are the datasets in your GEMS user dataset library. These Datasets are either temporary or permanent. Temporary Datasets are those which have been created during a GEMS session and which have not yet been saved. Permanent Datasets are those which have been saved and remain available from session to session. When GEMS procedures and models create new datasets, they are created as temporary. This is to help you conserve disk space. Temporary datasets are automatically deleted at the end of a GEMS session. If you do not need continuing access to a dataset, you should simply allow this deletion to take place in order to leave as much space available on the disk as possible. Whenever you exit from GEMS, you will be given a list of new datasets which were created during the current session. You will then be given an opportunity to save the new datasets, rescuing them from automatic and certain deletion.

The other type of dataset is the SYSTEM INSTALLED Dataset. These datasets are owned and maintained by the system and are available to all GEMS users. These datasets remain available from session to session and cannot be deleted or

modified by GEMS users. Datasets of this category contain data which has been found to be useful in performing environmental modeling by GEMS users. Many of these datasets are accessed automatically by various GEMS functions, or you can use them with various GEMS functions to create subsets which are useful for your own applications. For information relating to specific SYSTEM INSTALLED datasets please see Chapter 9.

Both User and SYSTEM INSTALLED datasets may be displayed by using the DATASET procedure under the OUTPUT function of the FILE MANAGEMENT operation. The DATASET procedure is documented in Chapter 5 (File Management) and you should review this section in order to gain a full understanding of the procedure.

An important concept which will help you understand the GEMS datasets is that of the data "group". Data groups are used to store information so that variables such as latitude and longitude which may have many records associated with them can be placed into one group and then the repeating records placed into another group with a link variable between them. This saves a great deal of disk space. If you do not realize which group a variable is in you can get quite confused when attempting to use a GEMS dataset (see Chapter 10 - Datasets).

The LIST DATASET command allows you to see all of your available datasets. Subcommands allow you to list selected datasets:

Enter a subcommand: TEMP, UON, UOFF, UALL, SON, SOFF, SALL, ALL, HELP, Return to end.
?

TEMP	All temporary datasets
UON	All on-line user datasets
UOFF	All off-line user datasets
UALL	All user datasets
SON	All System on-line datasets
SOFF	All system off-line datasets

SALL	All system datasets
ALL	All datasets
HELP	Brings Help

After you have identified a dataset of interest, the LIST VARIABLES command will allow you to list the variables in that dataset. If you are in a GEMS procedure that has specified a dataset, you will be shown the variables of that dataset. If you have not previously specified a dataset, GEMS will ask you to name one. The LIST DATASET command also has available several subcommands to allow you to specify which variables you would like to list.

Enter a subcommand: INT, REAL, CHAR, ALL, HELP, Return to end
?

INT	All integer variables
REAL	All real variables
CHAR	All character variables
ALL	All variables
HELP	Brings Help

The variables will be displayed by group.

The command DATASET will allow you to change the dataset for which you are listing the variables. Note: you can not do a LIST DATASET command from within LIST VARIABLES, or a LIST VARIABLES from within LIST DATASETS

Often throughout this user's guide, we will refer to your user dataset library. When we use this term we are referring to the data files and your MAP, GRAPH, REPORT, and TABLE datasets as well.

MAPS are created by the various GEMS mapping procedures, generally found in the GEOMAP procedures under the Geodata Handling operation. These datasets can only be displayed. No further manipulation can be performed on these data. Datasets may be reviewed by using the MAP procedure

under the OUTPUT function of the FILE MANAGEMENT operation.

GRAPHS are created by the GEMS graphics procedures found under the GRAPHICS operation on the main menu, as well as a number of specialized graphics procedures which are available for use with several of the models. These datasets can only be displayed. Further manipulation of the file is not allowed. These datasets may be reviewed by using the GRAPH procedure under the OUTPUT function of the FILE MANAGEMENT operation.

REPORT datasets store model output files. These files are placed within a structure which will allow you to better manage them. These datasets are usually text files which may be typed through use of the REPORT procedure under the OUTPUT function of the FILE MANAGEMENT operation. No further manipulation can be done to these datasets.

TABLES are generated by several GEMS procedures, for example, STATISTICS - Contingency tables. These datasets may be viewed by using the TABLE procedure under the OUTPUT function of the FILE MANAGEMENT operation.

The LIST MAPS, GRAPHS,... commands all act identically to the LIST DATASET command except for the absence of system maps, graphs, etc.

The LIST MAPS, et al. command offers you several useful subcommands:

Enter a subcommand: TEMP, ON, OFF, ALL, HELP, Return to end.
?

TEMP	All temporary maps, etc.
ON	All on-line maps, etc.
OFF	All off-line maps, etc.
ALL	All maps, etc.
HELP	Brings Help

1.4.4 Exiting GEMS

Now that you have mastered getting into GEMS, and have learned a little of the fun you can have with the system, it's time to learn how to leave GEMS.

If you are at any GEMS menu, you can simply type EXIT to leave the system. GEMS will then ask you to confirm your intention to leave the system:

Enter YES to confirm the EXIT command; type NO to restart GEMS...
?

If you haven't entered the EXIT command by mistake, confirm with YES and GEMS will check to see if you have created any new datasets while you were on the system. If you have, GEMS will ask if you want to SAVE them:

Temporary data created in the current session...

Datasets: ONE TWO THREE
Graphs: G1 G2
Maps: MAP1 MAP2

Press return to delete them, or enter "ALL" to save them, or enter only names of those temporary data to be saved.
?

Once you have taken care of these housekeeping chores, GEMS will return you to VMS, and you will see the VMS prompt.

$

Remember, if you are using GEMS for the first time, you must change your password before you log-off the VAX or you will not be allowed to log-on again!

Once at the VMS prompt, you can continue to use the system or log out. To log out, type LOGOUT

1.4.5 TGEMS

A special version of GEMS is used to make new procedures available for testing. This version of GEMS is called TGEMS. If you have set up the GEMS environment in your LOGIN.COM file, you can access this version by typing TGEMS. TGEMS is provided in order to make new procedures available to you as soon as possible during development. While procedures are not installed into TGEMS until they have been tested, these procedures are not always fully operational. Care should be taken when using any procedure in TGEMS. If you have a problem using TGEMS procedures, you should report it using the normal GEMS problems report method. The TGEMS menus may be different than those in GEMS, which are documented in this manual. This is because of new or updated procedures in TGEMS. The User's Guide is only updated when a procedure becomes operational and is installed into GEMS.

1.4.6 Reporting Problems

If a problem occurs while you are running GEMS or TGEMS, please let us know. It's easy, just use the PR procedure under the utilities menu, or send MAIL to the username GEX. It will help us understand your problem if you will remember to provide us with as much of the following information as possible.

- Date and Time the problem occured.
- Menu title of the menu on which you had problem.
- What happened. (Describe the problem as fully as possible.
- Error Messages (if any) that you received.

Remember, the more you can tell us about what happened, the better we can help. After you submit the problem report, we'll get back to you with an answer as soon as possible.

1.5 Commands Summary

GEMS has its own set of navigation and control commands segregated into four basic functional groups. The commands

discussed here are universally operable in GEMS. Additional commands may be functional in isolated sections of the system, and will be discussed as part of the discussion of those sections of GEMS. Note that the commands work as single word commands except where otherwise noted.

Help Commands obtain information about the system:

HELP	AUTOHELP	NOAUTOHELP
LIST	TUTOR	REVIEW
QUEUE	SHOW	

Navigation Commands allow you to move through the system:

BACK	END	CLEAR
EXIT	NEXT	GO

Operational Mode Commands support selection methods for data input:

MENU	PROMPT

Parameter Commands are used to change your operating parameters:

RESET	STORE	RESTORE

These commands, listed alphabetically, include the following. Unless specified, the command and the syntax are identical, that is to say, you enter the command exactly as it is written.

COMMAND	FUNCTION
AUTOHELP	Turns automatic help facility on, see "NOAUTOHELP" for a clue on how to turn it off.
BACK	Used to return to the previous prompt or menu screen.
CLEAR	Used to return to Main GEMS menu and restart the procedure selection operation.
EXIT	Used to leave GEMS. When EXITing, you must confirm your intention to leave the system. As you EXIT, GEMS will ask you if you want to SAVE any temporary GEMS datasets you have created during the

	current session. If you do not specifically save them, they will be deleted.
HELP	Used to call-up screens of system-help information. This help function is context sensitive. Entry of the HELP command affords you an opportunity to ask for specific help. If you enter HELP you will be given general help for your current screen. If you enter HELPn (where n is an option number on your current menu), you will be provided with help specific to that option.
LIST	LIST DATASETS lists the available user and system-installed GEMS datasets. LIST GRAPHS, LIST TABLES, LIST REPORTS, and LIST MAPS all list the available graphs, tables, reports, or maps in your user GEMS library. LIST VARIABLES lists the variables in the GEMS dataset you specify. LIST COMMANDS is used to list available commands. NOTE: There are two versions of this command, with the suffix "S" as in LIST MAPS, you will see a full list, without the "S", as in LIST MAP the list will be in abbreviated form.
MENU	Selects the Menu-Mode of system operation. This is the default mode, the alternate is PROMPT. The MENU command can also be used to refresh the screen if you are in menu mode.
NEXT	Calls the next parameter specification menu. Not applicable to navigational menus.
NO AUTOHELP	Turns off the AUTOHELP function.
PROMPT	Used to invoke the User Prompt mode of operation. Alternate mode is MENU.
QUEUE	Displays the current jobs in the system batch queue.

RESET	Resets parameter values to the default values on the original parameter editing menu.
RESTORE	RESTORE file-name, used to recall specification parameters saved with the STORE command.
REVIEW	Re-displays the current menu or previously entered parameter values in parameter editing menus.
SHOW	SHOW reference/parameter name displays current parameter values.
STORE	Used to store responses to parameter editing menus for later recall.
USERCMD	Displays available facility-defined commands.

Consult the sections of this manual which describe individual procedures for commands that are a part of individual programs, and are not detailed above.

1.6 Communication with PCs

As GEMS has grown, the number of GEMS users has also grown. This growth brought with it several problems, not the least of which was the prospect that the number of users would exceed the number which could be accommodated with the available processing power. The prospect loomed on the horizon that increasing numbers of GEMS users would result in a serious degradation in system response speed. An alternative was to utilize the increasing base of available PC processing power to absorb some of the new user demands.

1.6.1 PCGEMS

In order to meet the growing demands of EPA modelers for access to GEMS features, the Personal Computer (PC) version of the Graphical Exposure Modeling System (PCGEMS) has been adapted from the VAX version. PCGEMS has many of the same features as GEMS but is accessible at a local level on a desk top computer. PCGEMS features work both independently and in concert with their more powerful

counterparts on the mainframe. PCGEMS was created to meet two primary needs. First, PCGEMS reduces reliance on the VAX based GEMS by making GEMS capabilities independently available on a PC, and second, added capabilities allow a user to do preliminary work at the PC, send this information to GEMS on the VAX, and use the VAX based GEMS to perform the more demanding model runs.

PCGEMS is designed to reduce the reliance on the VAX based GEMS by providing the following independent capabilities:

1. Estimating the properties of chemicals such as octanol/water coefficient and boiling point, providing general chemical information of a chemical as well as information necessary to feed into the environmental models; it also estimates the fate of atmospheric pollutants in water and air.

2. Allowing use of environmental models which simulate the effects of chemicals released to the air, water and soil or groundwater. Plans are underway to include features to assess the risk to the ecology presented by chemicals and give estimates of the danger these chemicals may present to the environment.

In addition to these stand-alone features, PCGEMS was designed to work in concert with the GEMS on the EPA VAX Cluster at RTP. While PCGEMS has many useful features, it is not intended to accomplish all of the functions of its more powerful VAX namesake. However, PCGEMS has the capability to allow GEMS users to use PCs as work stations in preparation for using GEMS. In PCGEMS, you create datasets, build input files, and then transfer the information to GEMS on the VAX to access the more powerful GEMS capabilities. This PC-based pre-processing of data allows you to:

1. Work on a PC at times when the VAX is unavailable.
2. Reduce computer charges by shifting work to the PC.
3. Make use of powerful PC software packages in file preparation.

4. Keep local copies of vital data or programs.

The sections that follow provide more information on the communications capabilities built into PCGEMS to support this method of using GEMS. Refer to the PCGEMS User,s Guide for more details on PCGEMS

1.6.2 Communications methods to GEMS on the Cluster

There are a number of programs available to manage the transfer of files between the PC and the VAX. Kermit and CrossTalk are both tools in the standard EPA software distribution. PCPLOT is another useful terminal emulation package that makes a PC look like a Tektronix 4010 to the VAX. PCPLOT is typical of a number of similar products that allow you to use the graphics capabilities of GEMS operations on your personal computer.

Before you attempt to transfer a GEMS dataset to your PC, you must first use the G2VAX procedure to export the dataset from GEMS to a VAX file. You can not transfer a GEMS dataset directly from the VAX to your PC.

For each of the file transfer methodologies discussed below, a modem and the requisite software is assumed.

1.6.2.1 Communications with Kermit

Load Kermit by typing the command *KERMIT* then you will see the Kermit system prompt:

KERMIT-MS

If you do not have an initialization file for Kermit, then enter:

SET BAUD 1200 sets the communications speed to 1200 baud

SET PA EVEN sets the parity to even if calling on the Washington Direct Line.

SET PA NONE sets the parity to none if calling on the TYMNET network.

Now type:

CONNECT

Your PC will now act as a terminal emulator; that is to say your PC now looks like a regular computer terminal to the VAX. You can now logon to the VAX in the normal manner by entering ATDT(phone number) for a Hayes compatible modem. If you do not have a Hayes compatible modem, see your modem's user's guide for instructions on how to dial. Follow the instructions in section 1.3 for logging into the system.

To download a VAX file to the PC:

1. At the VMS prompt $, enter the command KERMIT to enter the Kermit file transfer program.

2. Enter SEND (filename) on the VAX. You may use wild cards to send multiple files with the same command (i.e., SEND *.FOR).

3. Enter ^ (^ = Control) and while holding down the control key enter C. This returns you to the MS-DOS KERMIT prompt where you will type RECEIVE (receive). KERMIT will then display an information screen to tell you what file it is currently receiving and how large it is. You may have to enter a CR (carriage return) to start the transfer.

4. When KERMIT has finished the transfer, just type C and you will be reconnected to the VAX.

To upload a PC file to the VAX:

1. Type KERMIT on the VAX, to enter the Kermit file transfer system. NOTE: If you are transferring a binary file (*.EXE), type SET FILE/TYPE=BINARY on the VAX.

2. Enter REC on the VAX. This will put the VAX into the receive mode.

3. Enter ^] (^ = control) and then enter C. This brings you back to MS-DOS KERMIT, from this command level, type SEND (filename). Wildcards are permitted, for example, SEND *.TXT. Kermit will display an information screen to tell you which file is being transferred, and the transfer progress. You may have to enter a CR to start the transfer.

4. When Kermit has finished the transfer, you can enter a C to be reconnected to the VAX.

1.6.2.2 Communications with CrossTalk

Load CrossTalk by typing the command CrossTalk. NOTE: A quick look at the status of CrossTalk is available by pressing the HOME key. The ESC key is used prior to entering any command.

If the CrossTalk software has already been set up for the VAX, enter the number corresponding to the appropriate command file. If none has been established:

Press the HOME key to see the status screen.
Press ESC to receive the COMMAND line.

Type the command prefix followed by the associated data to set the following parameters:

SPeed	2400	DUplex	Full
DAta	7	EMulate	VT-100 or VT-52
POrt	(1)*	MOde	Call
PArity	Even	CWait	Delay 5
STop	1	LWait	Delay 20

(* - Assumes modem on port 1)

Log on to the VAX:

Type *GO* at the Command prompt to dial the currently loaded phone number. If a new telephone number is to be selected, type [redacted] followed by the number, then type *GO*.

To download a VAX file to the PC:

1. Press the [redacted] key to go to command mode.

2. Type [redacted] followed by the drive designation for the drive where you want to write the incoming file, followed by the name you want to give to the file. (i.e., CAPTURE A:file.ext).

3. Type the information on the VAX (i.e., [redacted]) The file will then be transmitted from the VAX to the PC onto the specified drive and with the name you assigned.

4. Press the ESC key to go to command mode, and type [redacted] to save the file to disk.

To upload a file from the PC to the VAX:

1. First make certain that you are in the VAX line editor, EDIT, or in any system mode which will accept input data.

2. Press ESC to go into the command mode, and enter the command SEND, followed by the drive designation and file name of the file you wish to send to the VAX. (i.e.,[redacted]

3. Exit from EDIT by pressing the [redacted] (^ = control key) to save the file on the VAX.

To exit from XTALK, press the *ESC* key and enter [redacted]

1.6.2.3 Communications with PCPLOT

PCPLOT is just one of the many commonly used programs for Tektronix 4010 terminal emulation. We illustrate its use here only because GEMS users have successfully used it.

To upload or download files to the VAX using the PCPLOT program is simple with PCPLOT. Enter PCPLOT with the DOS level command PCPLOT. PCPLOT will first display a terminal type.

Make sure the terminal type is set it to either "1" or "2". If not, press ^T before proceeding. NOTE: A complete list of PCPLOT commands is available by pressing ^H.

Logon to the VAX by calling the appropriate telephone number with the ATDT (phone number) command.

To download a VAX file to the PC:

1. Depress the ALT key and simultaneously press the F key on your keyboard to name the destination file. This key combination will be indicated as ALT/F for the rest of this section. Include drive and full file specification in the name.

2. When ready to copy the VAX file, enter ALT/I to accept the host file. All terminal output will then be copied to the file you named.

3. To close the file, enter ALT/Q after the file has been downloaded.

To upload a PC file to the VAX:

1. You must be in the line editor, EDIT, in insert mode, or any other system mode which can accept input data.

2. Enter ALT/O to open the file on the PC.

3. Enter ALT/U to start uploading.

4. Enter ALT/R to close the file when uploading has been completed.

5. Exit from EDIT on the VAX to save the file.

When finished uploading or downloading using PCPLOT, enter ALT/X to exit to DOS command level.

Chapter 2 - Chemical Property Estimation

GEMS provides numerous programs to support physicochemical property estimation (we'll call it chemical property estimation). These estimation operations supply needed data on a wide range of chemicals. Table 2-1 lists these procedures and their function.

In this chapter we will first describe ways you can enter chemical structures. Then we will take a brief look at each of the GEMS estimation programs that are available to you. If you want to look at detailed information on specific procedures, you should consult the user's guides available for the individual program.

2.1 Describing Chemical Structures

Before we begin a discussion of the available property estimation techniques, you'll want to understand the methods available for describing a molecular structure for use by the GEMS programs. Two basic methods are available:

- SMILES, a linear notation method (GEMS preferred).
- SFILES, a two-dimensional molecule drawing method (Usable only with certain terminal types).

For many applications you don't need to describe the chemical structure; the chemical notations are already on file. GEMS contains a dataset of SMILES notations for over 19,000 chemicals. Section 2.2 will tell you how to use these pre-defined SMILES notations, and how to avoid having to enter the chemical notations.

2.1.1 SMILES Notation

The preferred method for describing and entering chemical structures into GEMS is called SMILES, the Simplified Molecular Input Line Entry System. SMILES was originally designed to bridge the gap between chemists and computers and is fairly easy to use. SMILES describes a chemical structure with a string of characters. These characters are comprised primarily of standard atomic symbols.

SMILES notations are hydrogen-suppressed, meaning that you do not need to indicate hydrogen atoms when the other atoms in the molecule are in their normal valence states. (Inorganic, organometallic, complexed, charged, and certain organic structures differ from simple organics in that it may be inappropriate to fill some atoms' normal valence states. However, most of these substances are not appropriate to use with GEMS estimation techniques.)

Nomenclature Rules

- In SMILES notation, all atoms are assumed to be covalently bonded unless you indicate otherwise. Aromatic bonds can be indicated by the use of the lower case letter c.
- Unless otherwise indicated, adjacent atoms are assumed to be bound by a single bond. Symbols for double and triple bonds are the equals sign (=) and the pound sign (#) respectively. Aromatic bonds can also be specified by the use of alternating single and double bonds.
- Branching is specified in SMILES notation by enclosing the branch in parentheses.
- Ring closure in cyclic structures is specified by mentally breaking any one bond in each ring and then assigning a unique number to each broken bond. When writing the SMILES notation for the structure, the atoms on either side of these bonds are followed with the number of the bond, as in c1ccccc1 for benzene.

- Additional hydrogens and formal charges are always specified inside brackets, as in [Na+].
- Ionic bonds are not specified in SMILES notation. Instead, the molecule is described as two separate structures separated by a period, as in [Na+].[C1-].

Table 2-1: GEMS Estimation Programs.

PROGRAM	FUNCTION	COMMENTS
SFILES	Store, modify, and retrieve a molecular structure diagram of a chemical.	Structure for SFILES can use SMILES notation.
SMIGET	Retrieves SMILES notation for any of 20K chemicals on file.	
DRAWSMI	Will draw any chemical structure from the SMILES notation.	
CLOGP3	Estimates octanol-water partition coefficient, Log P, using the Leo fragment constant method.	Input consists of molecular structure previously stored with SFILES, or direct SMILES entry.
CHEMEST	Estimates solubility in water, soil adsorption coefficient, bioconcentration factor, activity coefficient, boiling point, vapor pressure, rate of volatilization from water, and Henry's Law constant.	Property may be estimated by any of several methods. Uses its own commands.

PROGRAM	FUNCTION	COMMENTS
NEWCHEM	Enhanced version of CHEMEST estimates same properties plus acid dissociation constant and melting point.	
AUTOCHEM	All routines in CHEMEST but the rate of volatilization from water. Automatically selects the correct algorithm.	Methods and algorithms in NEWCHEM.
FAP	Estimates rate of atmospheric oxidation of a chemical by hydrogen radical and ozone. Rate of chemical deposition by adsorption on particles and dissolution in water droplets, if solubility and vapor pressure are known.	Combination of estimation methods. Input of chemical structure is via the chemical structure files created by SFILES.

SMILES Tips

Since the nomenclature rules are designed to provide a simple, consistent, and flexible language for specifying chemical structures, they enable even casual users to write correct SMILES notation. The easiest way to avoid trouble when you are writing SMILES notations is to remember four simple tips:

- There are 4 types of bonds in SMILES; single, double, triple, and aromatic. Ionic bonds are not specified directly.
- Bonds with both covalent and ionic character may be written either way, but pick one way and stick with it.
- Hydrogens for organic compounds are not normally specified in SMILES. Assume that hydrogens make

up the remainder of an atom's lowest normal valence consistent with explicit bond specification.

- SMILES will accept either aromatic or non-aromatic input specification, detect aromaticity, and internally convert the structure accordingly.

2.1.2 Molecular Structure Entry (SFILES)

The SFILES procedure was developed at Pennsylvania State University, and allows you to enter, store, modify, and retrieve a molecular structure diagram of a chemical through specific graphic display terminals. You must use this program if you wish to estimate the atmospheric half-life of a chemical using the FAP program (discussed in section 2.8). General Sciences Corporation has added a enhancing sub-program, SMILES, developed by EPA. SMILES allows users to enter a chemical structure through simple chemical structure notation. SMILES is generally used in conjunction with the CLOGP and AUTOCHEM procedures. The introduction of the SMILES enhancement to the SFILES program has virtually eliminated the need to use Textronix terminals to enter structures.

If you are going to use SFILES (required for FAP input, for example) you will need to have copies of the SFILES structure files in your LOGIN directory. To get copies of these files, use the GETFILES procedure (on the DIRMAN menu) to get a set of empty structure files. These files can be reused for all of your future needs. If you should accidentally delete your SFS files, the GETFILES procedure will let you get a new set of empty files. Chapter 8 contains more detailed information about directory management. If you are not familiar with directory management, You may want to read Chapter 8 to understand the DIRMAN and GETFILES procedures.

2.2 Retrieving Stored SMILES Notations (SMIGET)

As we mentioned, the Chemical Property Estimation routines in GEMS require information on a chemical's structure in order to

work. This structure can be specified by using either the SMILES or SFILES routines.

While use of the SMILES notation allows you the flexibility to use the estimation procedures on any terminal to which you have access, learning the syntax may take awhile. However, the use of SFILES to enter a structure by drawing it requires access to a Tektronix terminal. The SMIGET routine can be used to retrieve the SMILES notation for a chemical through use of the chemical's Chemical Abstracts Service (CAS) number. The SMILES notation is available to GEMS users for over 19,000 chemicals using SMIGET.

SMIGET also allows the creation of an output dataset in your GEMS user dataset library which stores the CAS number, SMILES notation and the chemical name for the chemical(s) with which you are working:

Table 2-2. SMIGET Output Dataset Elements

Variable Name	Description
SMILES	The chemical's SMILES notation
CASNO	The chemical's CAS number, stored as an integer
CHEMNAME	The chemical's name

This dataset can then be used as input to the GEMS estimation routines. The output created by SMIGET can also be exported from GEMS by using the File Management EXPORT function and downloaded to your PC for use in PCGEMS, or other desk-top applications. The SMIGET procedure is accessed through the Estimation menu or by the DAC [redacted].

2.2.1 SMIGET Input/Output Specification

When you access SMIGET, the parameter entry menu shown below appears Using this menu, you will identify the way in which

the procedure will get input information, and where the procedure will produce its output.

```
MENU: SMIGET Input/Output Specification

ref parmname    parameter description              value index
1. INMETHOD     input method specification         DIRECT
2. OUTPUT       name of the output DATASET file    BROWSE
3. TAG          Description tag field              *

Enter one or more combinations of : reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 2-1. SMIGET I/O Menu

You must identify three key parameters:

1. INMETHOD — SMIGET can accept input from either your terminal or from a DATASET in your GEMS User Dataset library. The use of GEMS dataset for input can be useful in retrieving the information on a number of chemicals. Specify DIRECT for terminal or DATASET to select the method you will be using. If you select the DATASET option then you must already have created the input dataset through use of the File Management procedures.

2. OUTPUT — This is the name of the output Dataset which SMIGET will create in your GEMS User Dataset library. If you answer BROWSE for this parameter then an output dataset will not be created. This allows you to use SMIGET to scan the information available in the database rather than always having to save it even when you don't want to. BROWSE is the default for this parameter. To create an output dataset you change the parameter to any legal GEMS dataset name.

3. TAG — Use the TAG parameter to give a description of up to 60 characters to the dataset entry which will be created for the output dataset specified by the OUTPUT parameter. If you do not set this parameter SMIGET will set it to "Created by SMIGET". When in BROWSE mode this parameter need not be set.

After completing this screen, enter NEXT. From here you will go to one of two screens depending on your previous entries. If you answered DATASET for the INMETHOD parameter then continue with section 2.2.2, however, if you answered DIRECT then you can skip that section and go straight to 2.2.3.

2.2.2 Input dataset designation

If you indicated DATASET in the previous menu, GEMS brings you to the input dataset designation menu. On this menu you will specify the input dataset which contains the CAS numbers to be used in extracting SMILES notations, and the variable in that dataset which contains these CAS numbers.

```
MENU: SMIGET1 Input dataset designation

ref parmname    parameter description                    valueindex
1. INPUT        Input Dataset name
2. CASVAR       Variable name which contains CAS num     CAS


Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS>
```

FIGURE 2-2. SMIGET Input Dataset Designation Screen

1. INPUT — You must name the input dataset which contains the CAS numbers of the chemicals for which SMIGET is to look up the SMILES notation. This dataset must have been created previous to your SMIGET session. It can contain any

number of variables, but one of them must represent the CAS numbers.

2. CASVAR — You must enter the variable name of the variable in the selected GEMS dataset which contains the chemical CAS number information. This variable can be either a character variable or an integer variable.

Once you have identified the file and the variable, use the NEXT command to run the SMIGET program and create the output dataset which you have specified. After you have created the new output dataset, you will be returned to the Estimation menu.

2.2.3 Retrieve SMILES notation

If you select the DIRECT option at the input/output menu, you will be prompted to enter the CAS numbers for which you want the SMILES notation:

Enter the CAS number to search for.
GEMS > [redacted]

The number we are using for illustration above could also have been entered without the dashes.

When you enter the CAS number SMIGET will first check to see that it is valid. If it is not a valid CAS number, you will see the following error message:

[ERR-404] You have entered an invalid CAS number

The **"Enter the CAS number to search for"** prompt will then be re-displayed.

If you have entered a valid CAS number, the database will be checked to see if the SMILES notation is available for the requested CAS. If it is you will see the following information on your screen:

The structure which was found is:
CAS NUMBER: 50-00-0

CHEMICAL NAME: Formaldehyde
SMILES NOTATION: O=C

The **"Enter the CAS number to search for"** prompt will then be re-displayed. If you are searching for more than one chemical, you may enter additional CAS numbers. When you have completed your search of the database, enter END to quit using SMIGET and return to the Estimation menu.

If the CAS number you entered could not be located in the database, then you will see:

[ERR-405] No data exists for the specified CAS number

You will remain in SMIGET, and may enter additional CAS numbers for which you would like to search the database. To end the procedure, type [illegible] and you will revert to the previous menu. Your output dataset will be created as you leave SMIGET.

2.3 Draw SMILES Structure

One of the easiest ways to validate your entry of a chemical structure is to see it actually drawn. This procedure allows you to do just that. You provide the SMILES notation of the chemical structure, and see it drawn on any Tektronix, or Tektronix compatible, graphics terminal.

Selection of option three on the Estimation menu invokes the DRAWSMI routine. You are then offered source and method options on the Input Option menu, shown below.

```
MENU: Input Option

ref parname      parameter description                value  Index
1  SOURCE        Source of Structure (SMILES/CAS)     SMILES
2  METHOD        Input Method (DIRECT/DATABASE)       DIRECT

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...]  or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 2-3. DRAWSMI Input Options Screen

After scanning the database, the system draws the structure. You can also use a CAS number as the source; the program will then search the chemical database for the SMILES notation before drawing the structure. Remember, DRAWSMI works only on Textronix compatible graphics terminals.

2.4 Octanol Water Partition Coefficient (CLOGP3)

The first estimation technique we will describe is called CLOGP3. It's used primarily to estimate a chemical's octanol water partition coefficient. The program name refers to this property as log P (for partioning); however, throughout these pages we will be referring to this property as log K_{ow}. This procedure was developed by the Pomona Medicinal Chemistry Project as part of the MedChem software. MedChem Software performs a wide variety of chemical information and modeling functions, available through a single, flexible access system called UDRIVE. GEMS provides you with access to the SMILES and the CLOGP3 capabilities through the UDRIVE front-end. CLOGP3 can draw the chemical structures, retrieve measured log K_{ow} values, and calculate log K_{ow} values. Before we discuss CLOGP3, we will have to discuss UDRIVE. For more detailed information about CLOGP3 and UDRIVE, you are urged to read the MedChem Procedures User's Manual.

2.4.1 The UDRIVE Access System

When you select CLOGP3 (option 2) from the Estimation menu, the UDRIVE access system first takes a moment to initialize and verify each installed module. This insures that all data files and algorithm control systems are correctly in place. If you are using an 80 column ASCII terminal, or the Tektronix 4014 graphics terminal, you will see a series of prompts that allow you to specify the input method, and then present you to a tabular output of the result. Remember, ASCII terminals can not draw pictures.

The prompts you will encounter on the 4014 or ASCII terminals will accomplish the same function we will discuss for the VT-100 terminal. If you are using a VT-100 or compatible terminal, your screen will show seven boxes around the top, right, bottom, and center of your screen. The center of your screen will display the initialization and verification checklist.

```
SMILES: UNDEFINED
NAME:  UNDEFINED

Unified Driver INITIALIZATION & VERIFICATION          CLOGP
                                                      Result:  N/A
1. MACSMI     initialization DEFAULT MACROS DEFINED
2. SMILES     verification SUCCESSFUL                 CMR
3. FRAGMENT   initialization COMPLETE                 Result   N/A
4. CLOGP3     verification SUCCESSFUL
5. CMR        verification NOT SUCCESSFUL             MEASURED
6. STARLIST   verification SUCCESSFUL                 LOGP   N/A
Comand? S
Input File: UNDEFINED          LINE 0                 ALSO
Keep File: UNDEFINED           LINE 0                 KNOWN
```

FIGURE 2-4. UDRIVE Initialization Screen

Each of the seven boxes on your UDRIVE screen is used for a specific input or output purpose. UDRIVE returns to this screen after every operation and updates these boxes after each calculation. The screen will be completely cleared when UDRIVE draws a chemical structure, but the screen will always return when these full-screen functions are completed. The UDRIVE screen gives you back the familiar **Command?** prompt.

Several commands direct UDRIVE. You can obtain a list of them by typing a question mark, ? at the **Command?** prompt. Only a single letter is required to execute each command. For example, to use the Q(uit) command to exit the program, you can enter just the Q. So that you will more clearly understand the commands and their function, we will include the whole command with the untyped portion of the command enclosed in parenthesis.

2.4.2 Using UDRIVE

There are two basic functions available under the GEMS installed UDRIVE. The first creates two-dimensional representations of molecules by using the P(icture) command. The second is the estimation and retrieval of measured log K_{ow} values. Both of these functions require the entry of chemical structure information. Direct entry of SMILES notation is the default method for specifying chemical structures. The task of entering and verifying structures is made simple with a range of input and output options.

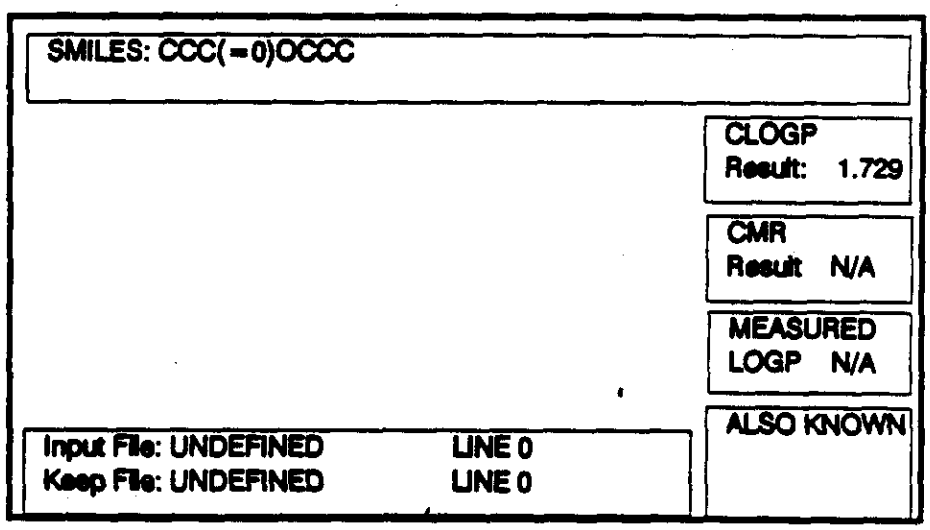

FIGURE 2-5. UDRIVE Screen with SMILES Notation

Typing S(tructure) at the Command? prompt on the UDRIVE screen allows you to directly enter a chemical structure. UDRIVE moves the cursor to the top window of the primary screen where you can enter your SMILES notation directly. The

following illustration shows the UDRIVE Screen with a SMILES notation already entered. You can use the *E(dit)* command to modify a SMILES notation.

Once you have entered a structure, UDRIVE allows you to save the SMILES notation to a file. Use the *S(ave)* command after entering the first chemical structure and UDRIVE will prompt you for the name of the output file. Subsequent use of this command will append each additional SMILES notation to the previously named file. Once you have entered the SMILES notation, you can also proceed to retrieving and estimating log K_{ow} values using the *C(logp)* command. This step is described more completely in the next section.

If you have created a file of SMILES notations (using SMIGET, for example) you can direct UDRIVE to read the contents of that file to draw the structures and/or to retrieve and estimate log K_{ow} values. These files could include files you have created using the GEMS *S(ave)* command and files created outside of GEMS. The *I(nput)* command causes UDRIVE to prompt you for the name of the SMILES notations input file. For files created outside of GEMS, be sure that they have .SMI or .MOL extensions. UDRIVE opens the file and reads the first SMILES notation. Enter the *N(ext)* command to proceed to the next chemical structure,. Pressing the return key without entering a file name returns you to the UDRIVE primary screen.

2.4.3 Introduction to Octanol Water Partition Coefficient (CLOGP3)

The MedChem software includes the CLOGP3 program. CLOGP3 provides a method for estimating the octanol/water partition coefficient (K_{ow}) for organic compounds. This procedure was developed by the Medicinal Chemistry Project at Pomona College, and utilizes Quantitative Structure Activity Relationships (QSAR) in an additive constitutive procedure to estimate the partition coefficient solely from the chemical structure. This discussion will provide a broad overview of CLOGP3, and is not intended to provide highly detailed instructions. If you need to know more about CLOGP3 and other MedChem software, you should read the MedChem documentation.

The fragment method of calculating log K_{ow} has proven to be valuable in many fields, including drug design and hazard assessment. However, manual calculations using the fragment approach require a great deal of instructions and become very lengthy for complex structures and, as a result, are error prone.

In principle, the measurement of the equilibrium concentrations of solute in the octanol and water phases, after shaking in a separator funnel, is very simple; and since good measured values are always to be preferred over calculated ones, it would seem that there should be little need for a procedure to calculate them. However, reliable shake-flask measurements are time-consuming and often difficult to make. Other measurement methods are also either time consuming or expensive, making the estimation of log K_{ow} a desirable alternative.

The Pomona MedChem Project saw these and other arguments as reason enough to develop a computerized method to calculate log K_{ow} from the chemical structure by using an additive constitutive procedure.

The first published method for calculating log K_{ow} from structure was based on a substitution procedure and was developed with substituent pi constants for aromatic rings in mind (the Rekker method). The Pomona method differs in the approach used to derive the actual working constants. The work of Drs. Jack Chou and Peter Jurs of Pennsylvania State University resulted in CLOGP, the forerunner of the present program.

Alkane carbons and hydrogens are the most fundamental fragments in the CLOGP3 system. An "Isolating Carbon" atom is a carbon which is not doubly or triply bonded to a hetero atom; and a fragment is any atom or group of atoms bounded by isolating carbon atoms.

2.4.4 Using CLOGP3

To properly perform its calculations, CLOGP3 needs to know the number and type of certain bonds in the solute structure. Once you have entered the chemical's SMILES notation (either directly or from a file), type [redacted] at the **Command?** prompt, and a search of the database of measured values is first conducted,

and then the program computes the log K_{ow}. The program then calculates the effect of all bonds within any fragment. When all calculations have been completed, CLOGP3 reports the estimated value. In addition, the program reports any measured value found in its database. You then receive the **Command?** prompt again, and you are free to estimate log K_{ow} for another chemical or leave the program.

Additional information on the operation of CLOGP3 and the specific application of the above mentioned factors on your computations can be found in the CLOGP3 user's guide published by MedChem, called the MedChem Procedures User's Manual. This users guide also includes a section with scores of examples.

2.5 Chemical Property Estimation (CHEMEST)

The CHEMEST program contains estimation routines for ten chemical properties. It was written by Arthur D. Little, Inc. for the USEPA Office of Pesticides and Toxic Substances and the U.S. Army Medical Bioengineering Research and Development Laboratory. The primary purpose of CHEMEST is to provide environmental scientists and managers with the capability of obtaining rapid estimates of these environmentally important properties of organic chemicals. The properties selected are those that are important in assessments of the environmental transport and fate of organic chemicals. The estimation methods are based upon one or more of the following: (1) empirically-derived regression equations between two or more properties; (2) theoretical equations with empirically-derived parameters; and/or (3) the summation of fragment and structural constants.

The current version of CHEMEST includes estimation methods for the following properties:

- Solubility in water
- Soil adsorption coefficient
- Bioconcentration factors for fish

- Activity coefficients
- Boiling point
- Vapor pressure
- Rate of volatilization from water
- Henry's Law Constant
- Melting point
- Acid Dissociation Constant

CHEMEST prompts you first for information on the chemical, the property, and the method of interest. Next a series of prompts will ask you for all required inputs such as the number and types of fragments present in the molecule. The property estimation loop is illustrated in Figure 2-6.

2.5.1 Using CHEMEST

The GEMS CHEMEST DAC provides access to the program. You can also enter this program by selecting CHEMEST from the GEMS Chemical Property Estimation Menu. You will first be asked a series of questions that establish the operating mode in which you will use CHEMEST:

"Do you want a brief description of CHEMEST?'

'CHEMEST operating modes currently available are...

1 Full prompted input
2 Limited prompted input (for new values only)
3 User control mode (without input display)
4 User control mode (with input display)

Type 1, 2, 3, or 4 to select desired operating mode:'

'Do you want to change the standard output display?'

'Type Y (Yes) to include the error summary in the display; type N (No) for deletion:'

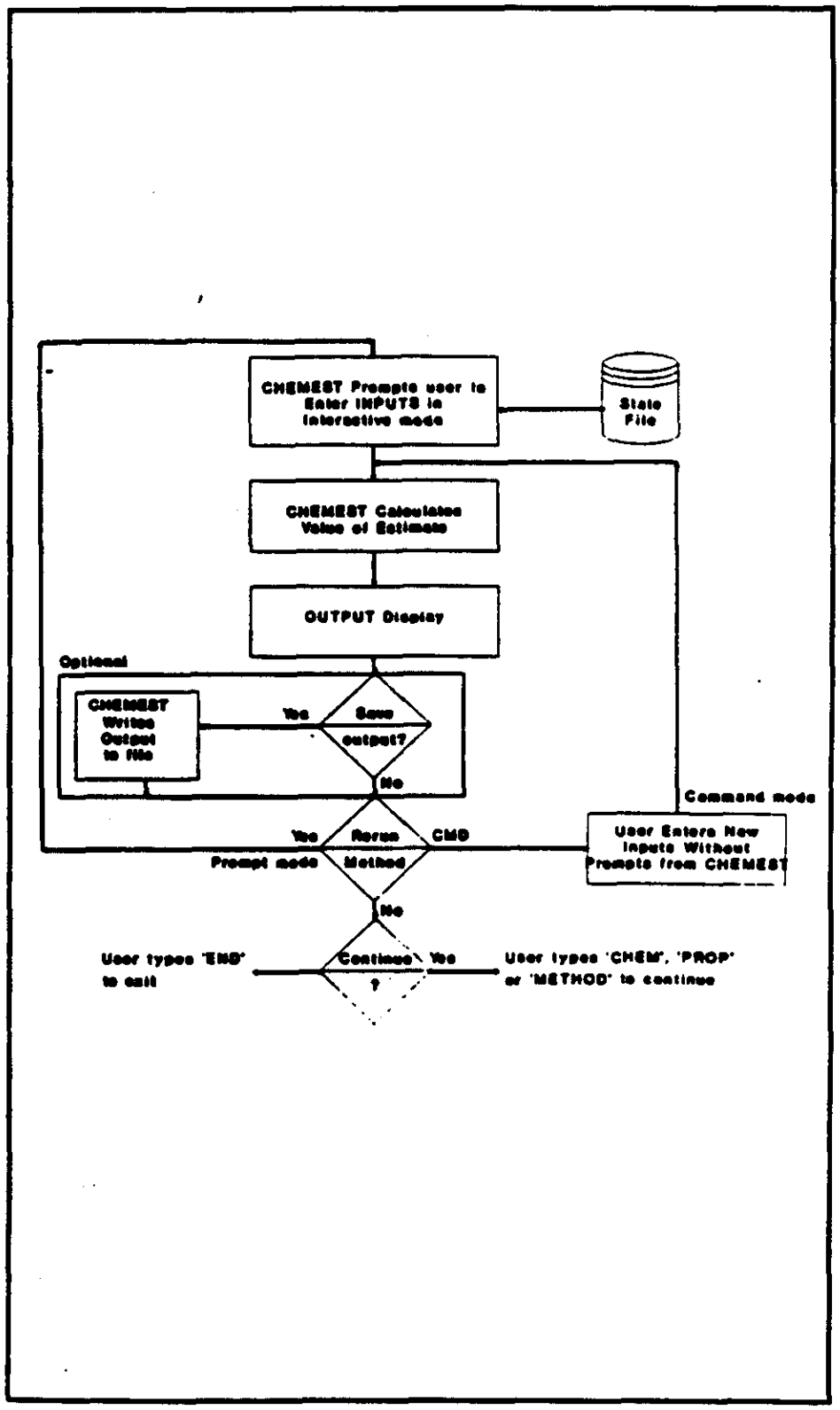

FIGURE 2-6. CHEMEST Property Estimation Loop

'Type Y (Yes) to include method identification in the display; type N (No) for deletion:'

'Type Y (Yes) to include key input values in the display; type N (No) for deletion:'

'Type Y (Yes) to include all other input values in the display; type N (No) for deletion:'

'Property estimation reports, to be displayed by CHEMEST at your terminal, may also be saved in an external file. The following choices are available -
1 No reports saved
2 Only selected reports saved
3 All reports saved

Option 1 is currently in effect; you must type 1, 2, or 3 to select the desired option:'

After you get through all of these set-up questions, you are ready to actually use CHEMEST.

There are several system commands used to direct CHEMEST, including:

HELP	Displays a brief guide to input procedures and list of system commands.
?	Displays a requested prompt (if available) that provides an explanation and/or example of the requested input. The original prompt message is then repeated.
END	Terminates the CHEMEST run.
STOP	Aborts current computation; go to select new chemical, property or method, or to terminate run.
RESTART	Aborts current run and restarts CHEMEST. Initiation procedures will have to be repeated.

CHEM	Specifies a new chemical will be identified for processing.
PROP	Specifies that a new property is to be selected.
METH	Specifies that a new method is to be selected.
PMT	Enter prompt mode of processing.
CMD	Enter command mode of processing.
NEW	CHEMEST only requests input of unspecified values. Switches from operating mode 1 to 2, or from 4 to 3.
OLD	CHEMEST will guide changes to previously specified values. Switches operating mode 2 to 1, or 3 to 4.
FILE	Allows you to change status of output file creation option.
FORMULA	Calls special routine to accept your entry of a chemical's empirical formula from which a molecular weight is calculated.
SUM	Special purpose command used to select re-entry of intermediate input for previously calculated run.
LAST	Special purpose command used for termination of multiple line input entries.
DISPLAY	Uses suffix to allow you to select options for display or estimated property and other outputs.
FLT	Allows you to test CHEMEST syntax for real input; follow command with value to be tested.
INT	Allows you to test CHEMEST syntax for integer input; follow this command with value to be tested.
NAM	Allows you to test CHEMEST syntax for alphanumeric input; follow command with value to be tested.

LA120	Only for use with the LA120 terminal to set normal spacing to 12 characters/inch.
VT100	Only for use with VT100 type terminals to set normal spacing to 12 characters/inch.

Additional information on these system commands and their use is available in the CHEMEST User's Guide.

Next, CHEMEST requires three items of information to go to work. The chemical name is specified first. The name is only for documentation purposes, but specification of a new chemical name will erase property values on parameters previously entered in the State File. After you enter the chemical name, you must specify the property to be estimated. You can change the property later by using the [redacted] command. Lastly, you specify the estimation method. To change the estimation method, enter the [redacted] command.

2.5.2 Estimating Property Values

CHEMEST can estimate ten different chemical properties using a number of different methods. These properties and the available methods are described below. Additional information about these properties can be found in the CHEMEST User's Guide.

Solubility in Water

Four methods are available to estimate solubility in water.

Method 1:	Solubility is estimated from the octanol-water partition coefficient via a regression equation. Three equations are used, one for organic acids, two for neutral compounds.
Method 2:	Solubility is estimated from atomic fragment constants and structural factors.
Method 3:	Solubility of polynuclear aromatic hydrocarbons is estimated from a

regression equation which is a function of N, the number of carbon atoms in the compound.

Method 4: Solubility is estimated from activity coefficients, which will usually be estimated values.

Specifics of these four methods, including method errors, propagated errors, required inputs, and flowcharts for coded methodologies, are presented in greater detail in section 4.1 of the *CHEMEST User's Guide.*

Soil Adsorption Coefficient

Four methods are available for the estimation of a soil (or sediment) adsorption coefficient, K_{oc}. All four are based on correlations between K_{oc} and one other property; the octanol-water partition coefficient, K_{ow} (method 1); water solubility, S (method 2); bioconcentration factor in fish, BCF (method 3); or parachor, P (method 4).

A total of 12 different regression equations are available. If measured values are available for all possible inputs the general order of preference amongst the methods is the same as the method number.

Additional information about the soil adsorption coefficient' including a flowchart of the coded methodology, input requirements, outputs, databases, and special features of the logic calculations is available in section 4.2 of the *CHEMEST User's Guide.*

Bioconcentration Factors for Fish

CHEMEST offers three methods for the estimation of a fish bioconcentration factor (BCF):

Method 1: Uses octanol-water partition coefficient (Kow)

Method 2: Uses water solubility (S)

Method 3: Uses soil adsorption coefficient per unit weight of soil organic carbon (K_{oc}).

The methods should be generally applicable to a wide range of chemical classes; deviations may be expected for chemicals that are easily metabolized.

All BCF values used in deriving these equations were from flow-through laboratory test under equilibrium conditions and ambient temperatures. No data from model ecosystem tests or static bioassays are included. All three regression equations used data from a variety of fish, and the BCF factors refer to the whole fish (wet weight); no correction based on fish lipid content is employed.

Additional information on bioconcentration factors for fish, including flowcharts of the coded methodology, input requirements, outputs, data bases, and special features can be found in section 4.3 of the *CHEMEST User's Guide*.

Activity Coefficients

Two methods are provided by CHEMEST to estimate the activity coefficients in binary systems. The only inputs required by CHEMEST to estimate activity coefficients are 1) information on the solute-solvent structures; 2) temperature of interest; and 3) mole fraction of solute (Method 2).

Method 1: The Pierotti Correlation, is restricted to the estimation of the infinite dilution activity coefficient. Furthermore, a limited number of solute-solvent systems and temperatures are available. In this method, the method error should be carefully considered.

The Pierotti Correlation method correlates the solute's infinite dilution activity coefficient with the number of carbon atoms in the solute and solvent. The correlation is restricted to a limited number of solute-solvent pairs, and can be applied over a temperature range of 10-100°C in most cases.

Method 2: The UNIFAC method is more general and can be used to estimate dilution activity as

a function of mole fraction and temperature. In this method, a group contribution method, interaction parameters have been calculated for a large number of structural fragments or groups. These interaction parameters, when incorporated into the UNIQUAC equation, allow you to estimate the activity coefficient of the solute and solvent as a function of mole fraction for a fairly general class of solute-solvent systems.

Additional information on activity coefficients and the two available methods, including flowcharts of the coded methodology, input requirements, outputs, data bases, and special features can be found in section 4.4 of the *CHEMEST User's Guide.*

Boiling Point

There are seven methods available in CHEMEST to estimate boiling point. In selecting the appropriate method, you should be mindful of several facts, namely: The only input required for any method is knowledge of the compounds structure.

Methods 1-3 are of general applicability and must be used for compounds with two or more functional groups. Methods 4 and 5 are preferable if the compound has only one functional group. Method 7 is the preferred method for saturated aliphatic hydrocarbons. Method 6 is preferred for hydrocarbons not covered by method 7.

The CHEMEST Users' Guide provides the following distinctions for the seven methods:

Method 1: Meissner	The boiling point is correlated with the parachor and molar refraction of the compound which are estimated via fragment constant methods. An adjustment is made for the compound class/type of interest.

Method 2: Lydersen, Forman, Thodos	The boiling point is estimated from the critical temperature. The constant of proportionality between boiling point and critical temperature is estimated via a fragment constant approach. Tc is estimated from the van der Waal's constants (a and b) which are also derived from the addition of fragment constants and structural factors.
Method 3: Miller	With Miller's method, boiling point is estimated from the critical temperature using the same proportionality constant as in Method 2. In this case, critical temperature is estimated from values of the critical pressure and critical volume which are derived from fragment constants.
Method 4: Ogata and Tsuchida	Boiling points are estimated in an empirical fashion via linear regression equations whose constants (slope and intercept) have been tabulated for various functional groups. The value of the independent variable in the regression equation is set by the length and structure of the base hydrocarbon radical.
Method 5: Somayajulu and Palit	This method estimates boiling points in an empirical fashion as well, using correlations with the atomic number sum of elements in the compound. The nonlinear correlation equation contains three constants (r, s, t) which have been tabulated for various chemical classes.
Method 6: Kinney	Boiling points are estimated in an empirical fashion via a linear regression equation with the boiling point number being obtained from fragment constants.
Method 7: Stiel and Thodos	An empirical correlation equation is used to relate the number of carbon atoms in the compound to the boiling point. For branched hydrocarbons, correction factors are introduced based on two

topological indices, the Wiener number and the polarity number.

Additional information on boiling point estimation, and the seven available methods, including flowcharts of the coded methodology, input requirements, outputs, data bases, and special features can be found in section 4.5 of the *CHEMEST User's Guide.*

Vapor Pressure

CHEMEST provides three methods for the estimation of vapor pressure.

Method 1: The vapor pressure is estimated through the use of the Antoine equation, but substitutes a modification of the Watson correlation describing the variation of the heat of vaporization with temperature. The required inputs are the normal boiling point and the identification of the compound class that best represents the polarity of the compound.

Method 2: In the second method, the vapor pressure equation is derived from a modification of the Watson correlation describing the variation of the heat of vaporization with temperature. The required inputs are the normal boiling point, the melting point for solids, and the identification of the compound class (as in method 1) that best represents the polarity of the compound. This method can be applied to both liquids and solids.

Method 3: The vapor pressure is estimated through use of the Antoine equation. The input parameters required are the normal boiling point and a structure dependent constant, K_F. If instead a boiling point at reduced pressure is available, the method allows you to estimate the vapor pressure at some other temperature.

Additional information on vapor pressure and the three available methods, including flowcharts of the coded methodology, input requirements, outputs, data bases, and special features can be found in section 4.6 of the *CHEMEST User's Guide.*

Rate of Volatilization from Water

There are two methods provided by CHEMEST to estimate the rate of volatilization from water. The first uses the Henry's Law Constant, while the second uses the oxygen regeneration rate. The method errors inherent in both have not been assessed quantitatively, so a factor of 5 has been assigned. In general, method 2 is probably preferred if a measured value of the chemical/oxygen ratio of volatilization rate constants is available. If such a rate is not available, method 1 should be used.

Method 1: Using Henry's Law Constant	In this method, the value of H is used to calculate the gas- and liquid-phase mass transfer coefficients; for chemicals with molecular weight 65, the water body depth, and wind and current velocities are also required. Options are included to allow you to include the effects of any added volatilization resistances associated with the transfer of the chemical from the bulk of the water to the interface, or the presence of surfactants on the interface.
Method 2: Using Oxygen Regeneration Rate	In this method, the two key inputs are the oxygen regeneration rate for the water body of concern, and the ratio of volatilization rates (preferably from laboratory measurements) for the chemical and oxygen. The entry of the oxygen regeneration rate may be avoided by entering values for the current velocity, water body depth, and (optional) river bed slope. The entry of the ratio of volatilization may also be avoided by entering a value for the chemicals diffusion coefficient in water.

Henry's Law Constant

CHEMEST provides three methods for the estimation of Henry's law constants for the air/water system. These methods are: (1) the use of the vapor pressure/water solubility ratio; (2) a structure-based (fragment constant) method developed by Hine and Mookerjee; and (3) the use of the activity coefficient X vapor pressure product. It is assumed that method 3 will most often be used in conjunction with estimated activity coefficients.

Method 1: Vapor Pressure/ Solubility Ratio	Henry's law constant may be equated to the vapor pressure/solubility ratio as H = Pvp/S. This equality is based on theory and assumes only that the solubility is low, that the total pressure is near one atmosphere, and that there is no association of molecules in the vapor phase. Method 1 is accurate whenever solubilities are less than about 1 mol/L. In select tests, estimates vs measured values differed by an average of only 0.5%.
Method 2: Hine and Mookerjee	Hine and Mookerjee derived a method of estimating H directly from structure. Two approaches are provided, one using a group contribution approach and one using a bond contribution approach. The group contribution approach should be used whenever possible since it has a lower method error. Methods 2 and 3 have been evaluated with a 77-compound test set of chemicals. Of these, 56 were either hydrocarbons (19) or contained only one functional group (37). The remaining 21 compounds contained two or more functional groups.
Method 3: Activity Coefficient x Vapor Pressure Product	The third method assumes only that the total system pressure is 1 atmosphere, that the fugacity coefficient is 1.0, and that the density of water is 0.9971. All are good assumptions for the purposes at hand. It is important to note that the

thermodynamic activity coefficient for a liquid refers to the liquid state and thus the value of vapor pressure product also refers to the liquid state. For solids, you must use the vapor pressure of the hypothetical supercooled liquid.

Additional information on estimation of Henry's Law Constant and the available methods, including flowcharts of the coded methodology, input requirements, outputs, data bases, and special features can be found in section 4.8 of the *CHEMEST User's Guide.*

Melting Point

CHEMEST provides two methods for the estimation of melting point, one developed by Lorenz and Herz, the other by Grain.

Method 1: Lorenz and Herz	This method uses a simple equation: Tm = 0.5839 Tb. Tb is the normal boiling point and both Tm and Tb are in degrees K. In addition to the normal boiling point, information with respect to whether the compound is aliphatic or not is required; this information is used to correct the calculated melting points. This method is generally applicable and requires only the normal boiling point as input.
Method 2: Grain	The Grain method is more complex requiring the input of an ID number for the compound, the liquid density, and the molecular weight, or the ID number, molecular weight, and structural information. This method yields a lower method error, but is restricted to compounds with C,H,N,O,S,Cl,Br,or I, unless the liquid density is known.

Additional information on estimation of the melting point and the two available methods, including flowcharts of the coded methodology, input requirements, outputs, data bases, and special features can be found in section 4.9 of the *CHEMEST User's Guide.*

Acid Dissociation Constant

CHEMEST provides two methods of estimating the acid and base dissociation constants. The Hammet or Taft equation, and the Delta pK_a method. For both methods, the only required inputs are those associated with a knowledge of the chemical's structure.

Method 1: The Hammet or Taft equation method has a method error of 0.2. The estimate is based on specific parent structure. Approximately 200 parent structures are available, with about 600 substituents.

Method 2: The Delta pKa method has a method error of 0.5 with the estimate based on generic parent compound class. Use of method 2 is limited to aliphatic carboxylic acids and aliphatic amines.

Additional information on estimation of the acid and base dissociation constants, and the two available methods, including flowcharts of the coded methodology, input requirements, outputs, data bases, and special features can be found in section 4.10 of the *CHEMEST User's Guide.*

2.6 New Chemical Property Estimation (NEWCHEM)

NEWCHEM is the upgraded version of CHEMEST. Both can be used. The older version (CHEMEST) follows closely the handbook for water solubility estimation methods, while the newer version uses newer equations and does not follow the handbook. You should be aware that the results from the same water solubility estimation method number run under both versions will yield different results. With this notable exception, NEWCHEM performs virtually the same as CHEMEST.

2.7 Automatic Chemical Property Estimation (AUTOCHEM)

The AUTOCHEM procedure, was developed for EPA by General Sciences Corporation. It is an automated chemical property estimation system.

AUTOCHEM provides estimation routines for octanol/water partition coefficient ($LOGK_{ow}$), melting point (MP), Henry's Law Constant (HC), bioconcentration factor (BCF), and adsorption coefficient for soils and sediments (K_{OC}). $LOGK_{ow}$ is estimated using Leo's CLOGP3 program and other routines are based on the methods and algorithms used in CHEMEST. Only those methods having general applicability have been implemented in AUTOCHEM.

All you need to know to use AUTOCHEM is the chemical structure and how to enter it into GEMS using either SMILES notation or SFILES. In fact, once a chemical has been entered in SFILES or its SMILES is saved in a GEMS dataset, AUTOCHEM can be used by a non-chemist. AUTOCHEM can also retrieve SMILES notations from a file of more than 19,000 chemicals when you enter the chemical's CAS number. AUTOCHEM is capable of estimating all of the commonly used chemical properties without any further action on your part.

2.7.1 AUTOCHEM Procedures

The command AUTOCHEM will allow you access to AUTOCHEM. The welcome to AUTOCHEM is the first thing you will see and then you will be asked if you want to save the summary table from this session as a GEMS dataset, if so, under what name, and the input method and option:

AUTOCHEM CHEMICAL PROPERTY ESTIMATION SYSTEM

(AUTOCHEM)

VERSION 1.90 (8-8-1988)

(AUTOCHEM NEWS - Any news about the system will be displayed immediately after the version identification)

Save Summary Table as GEMS Dataset (YES, NO)
Default = NO
GEMS > YES

Enter GEMS Dataset name (up to 20 characters)
GEMS > TESTAUTO

Enter the structure input method (SMILES, CAS, or SFILES)
Default = SMILES
GEMS > SMILES

The SMILES input method, provides you with two options. You may enter the SMILES notation directly or have the SMILES notation read by AUTOCHEM from an existing structure file.

Enter input option (DIRECT or GEMS)
Default = DIRECT
GEMS >

If you had entered DIRECT as your desired input method, you would be asked to provide the SMILES notation. For purposes of discussion, we will use a typical SMILES notation.

Enter SMILES Notation
GEMS > [illegible]

If you choose the GEMS option, enter GEMS at the GEMS prompt, and the SMILES Input Option menu appears:

```
MENU: SMILES Input Option

ref parmname   parameter description          value  index
1. DATASET     SMILES dataset name
2. SMILEVAR    SMILES notation variable name
3. CASVAR      SMILES notation variable name
4. NAMEVAR     Chemical name variable name
5. RANGE       Read a range of structures

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command:
HELP,NEXT,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 2-7. SMILES Input Option Menu

This option allows you to enter the name of an existing SMILES notation file, add new structures to existing files; or create a new SMILES notation file.

The second input method is to enter the CAS number.

Enter the structure input method (SMILES, CAS, or SFILES)
Default = SMILES
GEMS > CAS

Then you enter the CAS number.

Enter CAS Number without dashes
GEMS > 87-64

If the SMILES for the entered CAS number is invalid or not found, you will be informed and given an opportunity to enter another CAS number. If you need to, you may also back up at this point, using the BACK command, to select a different input method.

The third input method is to select SFILES. The structural information about the chemical is obtained by AUTOCHEM from existing structure files. When you choose the SFILES input method, you then specify the SFILES Information:

Read a range of structures (YES, NO)
Default = NO
GEMS >

Enter YES or NO to this prompt. Next, a menu will appear that will vary slightly depending on what you answer. The menu that will appear if you answer YES is shown below.

```
MENU: SFILES INFORMATION

ref parmname    parameter description              parameter value  index
1. BEGDAN       SFILES beginning direct access number  1
2. ENDDAN       SFILES ending direct access number     1
2. DIRECTORY    Directory for structured files         user12$disk:[imr]
3. TYPE         File type for SFILES                   SFS

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command:
HELP,NEXT,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 2-9. SFILES Information Menu

The illustrative menu shown above has already been completed. If you answered NO to the previous prompt, the menu above will ask you to enter only a single Direct Access Number. Once this has been accomplished, GEMS will advance to the Edit Property Values Menu. Next you are allowed to edit the property values that will affect your estimation run.

AUTOCHEM will look in a couple of the databases for measured values for any of the first five properties and chemical name. If any of the property values are found in the databases, they will be displayed on the menu under parameter value(e.g., 1.99 below). At this menu, you may enter any known property values using the format shown at the bottom of your menu (ref1 value1, ...). The asterisk entered in the parameter value denotes an unknown value, properties with asterisks will be estimated by AUTOCHEM.

```
MENU: Edit Property Values & Commands EDIT, NEWS, DISPLAY

ref parmname    parameter description                  value  index
1.  LOGKow      Octanol water partition coefficient    1.99
2.  BP          Boiling point                          *
3.  MP          Melting point                          *
4.  WS          Water solubility                       *
5.  VP          Vapor pressure                         *
6.  HENRY       Henry's law constan                    *
7.  BCF         Bio-Concentration Factor               *
8.  KOC         Adsorption coefficient                 *
9.  STATE       Physical state                         UNKNOWN
10. BPUNIT      Boiling Point Unit                     CELSIUS
11. VPTEMP      Vapor pressure temperature             25
Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command:
HELP,NEXT,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 2-8. Edit Property Values Menu

If the chemical name is not found, then you will be prompted to enter the name. In our example, the name was found and this prompt was not shown. There are three specific AUTOCHEM commands available at all prompts. To use these commands, enter the "EDIT" command at any prompt and the Switch Menu appears:

```
MENU: SWITCH

ref parmname    parameter description                  value  index
1.  MTDERR      Use the method error routine           NO
2.  PARMTR      List parameters                        YES
3.  WAIT        Pause between property estimation      NO
4.  EDTMP       Change Melting Point parameter         NO
5.  EDTBP       Change Boiling Point parameter         NO
6.  EDTVP       Change Vapor Pressure parameter        NO

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command:
HELP,NEXT,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 2-11. Switch Menu

Finally, before you can run your estimation, you must define which properties you want AUTOCHEM to estimate:

```
MENU: Chemical property estimation and commands EDIT, HELP, NEWS

ref parname     parameter description              value index
1. CLOGP        Estimate Log Kow using CLOGP3      NO
2. BP           Estimate Boiling Point             YES
3. MP           Estimate Melting Point             YES
4. WS           Estimate Water Solubility          YES
5. VP           Estimate Vapor Point               YES
6. HENRY        Estimate Henry's Law Constant      YES
7. BCF          Estimate Bio-Concentration Factor  YES
8. KOC          Estimate Adsorption Coefficient    YES

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 2-9. AUTOCHEM Menu

When you enter NEXT AUTOCHEM then begins to produce output in response to the parameters you have specified. For this illustration, we will examine the results for the example "C = CC = C" which we input earlier.

****

Mol Wt = 54.09 Formula = C4H6

****

Boiling Point Estimation

Boiling Point

Method	BP(K)	BP(C)	N	E.F.
Meissner	277.9	4.7	0	Not Used
Miller	270.4	-2.6	0	N/A

Melting Point Estimation

Melting Point

Method	MP(C)	MP(K)
Grain and Lyman	-105.04	168.12

Water solubility at 25.0C

EQN	WS(mg/1)	N	E.F.
2	1848.49	0	0.00
2C	2687.30	0	0.00
3	223.47	0	00.0
4	1690.51	0	00.0
12N	1528.79	0	00.0

AUTOCHEM will use equation 12N results for other estimations. Water Solubility is estimated by using estimated melting point.

Vapor Pressure Estimation

Vapor Pressure at 25.0(C)

Method	Kf	VP(mm)	N	E.F.
Antoine	1.010	1.534E+03	0	0.00

Estimated or entered BP used = 277.9(K)
For gases only Antoine method is used.

Henry's Law Constant at 25.0 C (atm m3/mol)

VP(mm)	WS(mol/L)	HLC
1.000E+00	2.826E-02	3,892E-02

Estimation error may be high for the following reason: Water Solubility > 1 Mol/L, HLC may be higher than 3.892E-02

Note: For gases Vapor Pressure used = 760 mm Hg

Bio Concentration Factor

Method	LogKow	BCF
Kow	1.990	19.2

Adsorption Coefficient (Koc)

Eqn	Method	LogKow	Koc
4-10	Kow	1.99	60.3

The above output product is occasionally more detailed than you really need. If it is, you can review summary tables that give you the over-view of this long list of information.

Press RETURN to see summary table
GEMS > {carriage return}

The following summary table appears in response.

Physico-chemical Property	Estimated or Entered Values	Method and/or Equation *
LOGKOW	1.990E+00	Measured *+
Boiling Point	4.753E+02 (c)	Meissner
Melting Point	-1.050E+02 (c)	Grain and Lyman
Water Solubility	1.529E+03 mg/L 25c	12N
Vapor Pressure	1.534E+03mmHg 25c	Antoine
Henry's Constant	3.892E-02 atm m3/mol	Method 1 +*
BCF	1.916E+01	Kow, (method 1)
Adsorption Coeff.	6.026E+01	Kow, Eqn. 4-10

* Equations used and the selection rationale for the most reliable value are described in the AUTOCHEM manual.
*+ From the MEDCHEM database.
+* For gases VP used = 760 mm Hg

FIGURE 2-10. AUTOCHEM Output Summary Table

Details on specific estimations and the manner in which they function within AUTOCHEM, and additional information about other AUTOCHEM functions, are contained in the *Automated Procedures for Physicochemical Property Estimation* (GSC-TR8664). This manual is to be replaced in the near future with an updated version.

2.8 Fate of Atmospheric Pollutants (FAP)

FAP was designed by Heicklen Associates as a method for estimating the rate of atmospheric oxidation of a chemical by hydroxyl radical and ozone from structural information alone. If the solubility and vapor pressure of the chemical are known, FAP can also estimate the rate of deposition of the chemical by adsorption on particles and dissolution in water droplets. The FAP program is accessible from the GEMS Estimation menu.

You can use only upper case letters when entering information for FAP.

The FAP model requires the SFILES input of the chemical structure for the model calculations. FAP estimates atmospheric reaction rates by identifying structural fragments within the molecule which are susceptible to chemical reaction. Associated with each of these fragments are constants which are proportional to the chemical reaction rate. The total reaction rate for a chemical is obtained by summing the contribution to the reaction rate made by each of the identified fragments. FAP calculates second order reaction rates, dependent upon the concentration of one of the reactants. The FAP model then converts these reaction rates to first order, for each of the following processes: abstraction of hydrogen, addition of hydroxyl radical to double bonds, addition of hydroxyl radical to aromatic bonds, reaction with ozone, heterogeneous removal, surface lifetime, and lifetime in the troposphere, which is 15 kilometers above the earth's surface.

Method 1: Abstraction of Hydrogen	The method use by FAP to estimate the rate of hydrogen abstraction by hydroxyl radicals (developed by Heicklen) uses bond dissociation enthalpies for each carbon to hydrogen bond. The sum of all the carbon to hydrogen bonds is used in calculating the reaction rate for the entire molecule. This method is unique in that it is the only estimation method in the FAP model which allows the temperature dependence of the rate to be determined.
Method 2: Addition to Double Bonds	The method which is used to estimate the rates of hydroxyl radical addition reactions is a modified version of that developed by Hendry and Kenley. The estimated rate of attack of OH on a double bond. The total halogen substituent correction factor is the product of the halogen substituent correction factors for each of the halogens attached to the double bond. The principal difference in this method is that it does not recognize

a difference in rate for cis and trans dialkyl substituted olefins, rather an average value of the partial rate constants for the two isomers is used. A modification of Method 2 can also be used for addition to triple bonds.

Method 3: Addition to Aromatic Bonds

Estimates of the rate of addition of hydroxyl radicals to aromatic rings are also calculated using the Hendry and Kenley equation. For addition to aromatic bonds substituent, halogens are treated in a fashion similar to that for addition to double bonds. However, for addition of hydroxyl radicals to aromatic bonds the halogen atoms F, Cl, and Br all have the same basic correction factor of 0.3. The correction factor for multiple halogenation is obtained by multiplying together the correction factors for each of the halogen atoms.

Method 4: Reaction with Ozone

The rate of reaction of ozone with organic compounds is also estimated using the same equation. The list of partial rate constants used in the FAP model is again a modified version of the correlations of Hendry and Kenley. Ozone does not react with alkanes or with aromatic rings. In general the rate of reaction of a compound with ozone is much slower than the rate of reaction with hydroxyl radical. The actual rate of reaction is dependent on the concentration of ozone, which may vary significantly in urban environments.

Method 5: Heterogeneous Removal

The heterogeneous removal of a compound is calculated by the FAP model according to a theory developed by Heicklen. This calculation routine requires that you know the normal vapor pressure and solubility of the compound. If these values are not known, a useful rate of heterogeneous removal can not be estimated. You can instruct FAP not to enter the heterogeneous removal routine. This routine is based on the assumption that molecules of the compound can be permanently removed from the air by adsorption onto aerosol particles and dissolution into water droplets. These particulates and water droplets in turn are removed from air by natural processes with a lifetime in air of about 4,000,000 seconds.

Method 6: Surface Lifetime

The surface lifetime of a chemical is estimated by summing the second order rates of removal calculated by the routines described above. These second order rates are converted to pseudo-first order rates by multiplying the hydroxyl reaction rates by the hydroxyl radical concentration of 800,000 molecules per cubic centimeter and the ozone reaction rates by the ozone concentration of 600,000,000,000 molecules per cubic centimeter. The lifetime of a reacting species is calculated as the reciprocal of the total pseudo-first order rate constant. The lifetime (or half life) calculated by this routines assumes that the rate of removal is fast enough that the chemical does not become globally mixed in the troposphere.

Method 7: Lifetime at altitude of 15 Kilometers	For a chemical with a long lifetime, atmospheric mixing processes distribute the chemical on a global basis throughout the troposphere. The removal processes described above are limited by their many assumptions for estimating near surface removal. For the special case where hydrogen abstraction by hydroxyl radical dominates the removal process and thus the temperature dependence of the removal process is known, the lifetime at 15 km is recommended. This routine calculates the ratio of the global lifetime to the surface lifetime. The global lifetime and the global half life may be determined by multiplying this ration by the surface lifetime and half life.

2.9.2 Running FAP

The execution of FAP is simple and can be accomplished with the use of the commands listed below:

CALC	This command tells the model to calculate reaction rates and lifetimes. Calculations may be performed either for structures in a worklist, or if the worklist is empty, the model will prompt for the direct access number of the chemical.
DIRS	This command prints a list of the commands.
DONE	Used to exit FAP.
DRAW	Usable only with a Tektronix graphics terminal, this command draws the structure.
HELP	Help lists all of the available commands and short explanations of their use.

HETR	This command allows you to turn the heterogeneous removal routine off and on.
LIST	Allows you to obtain a list of the structure DAN's labels and names of structures in the structure file. This command prompts you to enter a range of DAN's to be listed.
NOPU	Instructs FAP not to pause between calculations for structures in a worklist.
PAUS	Causes FAP to pause, and allow you to inspect the results, between calculations.
PRIN	Changes the amount of output you receive as a result of model calculations. The default print level is 1 (one).
TEKT	Informs FAP that you are using a Tektronix graphic terminal.
TERM	Informs FAP that you are not on a Tektronix terminal.
WKLS	Allows you to create and delete a calculation worklist. WKLS has its own subset of commands:

- ADD allows you to add a sequence of DANs to the worklist,
- DELE allows you to delete DANs or a range of DANs,
- SORT allows the entire worklist to be sorted in ascending order,
- NEG negates all entries in the worklist,
- PRNT directs the printing of DAN values in the worklist to the printer,

- CHEC lists the DAN values on the terminal screen,
- MAIN returns to the main FAP prompt level.

Chapter 3 - Environmental Modeling

GEMS provides access to a host of modeling tools designed to enable the user to simulate the migration and transformation of a chemical in the environment, either in a single medium (e.g. surface water) or as it moves through all media. Many of these models were designed as stand-alone systems and later integrated into the GEMS environment. A number of the models have menu-type prompting provided by GEMS, while the rest use the prompts originally designed within the model.

Table 3-1 below provides a list of the models available in GEMS, organized by media. For each model, the table shows whether the model output is stored as GEMS datasets (reviewable using the File Management function in GEMS) and which models have stored data available to the user. The table also shows the type of prompting employed by the model; GEMS prompting indicates the development of a model interface to allow menu prompting. Note that while GEMS provides these interfaces, with few exceptions, the models themselves have not been modified.

In general, there are four steps involved in using a GEMS model: data input, running the model, viewing output tables, and creating output graphics. The sections that follow briefly describe each of the GEMS models, the input information they require, and the resulting model output. For more detailed information about understanding and using the models, we include references to GEMS-specific users' manuals, and model documentation written by the model developers.

TABLE 3-1. Models Available in GEMS

MODEL	OUTPUT TO FILE MGMNT	STORD INPUT DATA AVAIL	GEMS PROMPT	MODEL PROMPT
AIR				
GAMS	YES	YES	YES	
– ISCLT	YES	YES	YES	
– TOXBOX	YES	YES	YES	
INPUFF	YES	YES	YES	
PTPLU	YES		YES	
BOXMOD		YES		YES
SOIL				
SESOIL	YES	YES	YES	
PRZM				YES
GROUNDWATER				
AT123D	YES		YES	
SWIP	YES		YES	
SURFACE WATER				
EXAMS II	YES	YES	YES	YES
MULTIMEDIA				
ENPART			YES	
TOX-SCREEN	YES	YES	YES	
MICROBE-SCREEN	YES	YES	YES	
UTM-TOX				

3.1 Air Models (AIR)

GEMS provides you with four different models to estimate concentrations of chemicals released into the air, including:

- GAMS
- INPUFF
- PTPLU
- BOXMOD

GAMS is an integrated atmospheric modeling system that can be used to estimate annual average concentrations, annual exposure, and lifetime and annual incidence of excess cases of cancer. INPUFF is a Gaussian plume, integrated puff model capable of addressing the accidental release of a substance over several minutes, or of modeling the more typical continuous plume from a stack. PTPLU estimates the location of the maximum short-term (one hour) atmospheric concentration from a single source as a function of stability and wind speed. PTPLU can also be used to estimate short-term concentrations as a function of downward distance. BOXMOD is a steady-state simple atmospheric area source model applicable to regions containing many diffuse emission sources within its boundaries, such as in an urban area. BOXMOD calculates a single annual average concentration applicable to the entire region based on a uniform area emission rate.

The user selects from these models depending on the type of estimates that are required and on the make up of the release scenario. In general, BOXMOD and PTPLU would be used in a detailed screening step of an analysis and INPUFF and GAMS would be used when the detailed screening indicates that a more refined analysis is required.

3.1.1 GAMS

The GEMS Atmospheric Modeling Subsystem (GAMS) incorporates the Industrial Source Complex Long-Term (ISCLT) model and the TOXBOX urban area source model. GAMS allows the user to build model input data files for these models, and also allows model output to be combined with data on the distribution of surrounding populations to calculate exposures as well as environmental concentrations of the released chemical. GAMS allows you to examine overlapping exposures by integrating the atmospheric concentration estimates of the two models with a population distribution data base to estimate exposure and risk.

GAMS can consider up to twenty source categories (e.g. chemical production or major uses) with up to nine emission type (e.g. process or storage) entries within each category for

an unlimited number of source locations for ISCLT modeling. TOXBOX, capable of handling up to 10 source categories, is implemented in GAMS to estimate annual average concentrations for all U.S. urban populations. The urban population comprises all persons living in the 366 urbanized areas (UAs) and in the 3827 places of 2,500 or more inhabitants outside urbanized areas. Land areas used for the urban populations were obtained from a 1980 Bureau of the Census publication (Census, 1983).

The concentration estimates generated by GAMS models are stored at the Census block group and enumeration district (BG/ED) geographic level. The population database used in GAMS was created from the proprietary Master Area Reference File (MARF). MARF is a datafile released by the Bureau of the Census to which Donnelly Marketing Information Services has added the latitude-longitude coordinates for all BG/EDs. Both 1980 Census population figures and 1983 population estimates provided by Donnelly (1983) are available within GAMS. By assigning concentration estimates at this detailed population distribution level, GAMS accounts for increments of concentration from any number of sources that may impact an individual BG/ED. This tracking avoids multiple counting of populations when sources are close enough to one another that surrounding BG/EDs are impacted by more than one source.

Exposure and risk calculations can be performed by GAMS for each BG/ED population by source category and emission type from the ISCLT results, by source category from the TOXBOX results, and across all source categories from the overlapping results from both models. Excess risk over background is estimated within GAMS by using a potency slope factor, a unit risk factor, a multi-stage model, or a one-hit model.

You may enter a potency slope factor (q*) in the units of $(mg/kg/day)^{-1}$ or a unit risk factor in the units of $(ug/m^3)^{-1}$. If excess risk is estimated using a multi-stage or one-hit dose-response model, several parameters such as the daily inhalation rate, body weight, q parameters, and number of stages are required. The user is referred to the *GAMS User's*

Guide (GSCTR-32-88-017, 1988) for complete details on how the exposure and risk calculations are performed.

There are six procedures in the GAMS menu of GEMS: (1) Set up Atmospheric Modeling (GAMSATMOSET), (2) Run Atmospheric Modeling (GAMSATMORUN), (3) Exposure and Risk Estimation (GAMSERE), (4) GAMS Graphics (GAMSGRAPH), (5) GAMS Utilities (GAMSUTIL), and (6) Clean up a GAMS Study (GAMSCLEAN).

The first procedure, GAMSATMOSET, is used to set up the modeling scenario of your study. The second procedure, GAMSATMORUN, is used to run the atmospheric modeling study previously set up. The third procedure, GAMSERE, is used to set up and run exposure and risk estimations on your study.

The fourth procedure on the primary GAMS menu, GAMSGRAPH, is used to obtain graphics from the results of the ISCLT modeling within GAMS. Three different types of graphs can be generated: BG/ED, ISOPLETH, and OVERLAY. The BG/ED graph displays the distribution of Block Group/Enumeration District (BG/ED) population centroids around the modeled site. The ISOPLETH graph displays closed isopleths of ISCLT estimated concentrations. The OVERLAY graph displays the isopleths of concentration and BG/ED population centroids.

The fifth procedure, GAMSUTIL, only pertains to TOXBOX modeling. The utilities are used to modify or create state and county emission rate files that are used as input to the TOXBOX model.

The sixth procedure, GAMSCLEAN, is used to delete various files that are associated with a study you are working on. You will usually use this procedure when you decide that your study should be rerun. Consult the chapters on GAMSCLEAN and system considerations in the *GAMS User's Guide* for additional information on this procedure.

3.1.1.1 ISCLT

ISCLT was developed for the EPA, Office of Air Quality Planning and Standards. The ISCLT model code corresponds to that found in UNAMAP Version 6. ISCLT combines and enhances various dispersion model algorithms into a computer program that can be used to assess the air quality impact of emissions from the wide variety of sources associated with an industrial source complex.

ISCLT may be used as either a screening or a detailed level model to estimate atmospheric concentration or deposition values. It is possible to estimate populations exposed to various levels of pollutant concentration and the population exposure through the integration of ISCLT within GAMS. The ISCLT model is a sector-averaged model that extends and combines basic features of the Air Quality Display Model (AQDM) and the Climatological Dispersion Model (CDM). The long-term model uses statistical wind summaries to calculate annual ground-level concentration or deposition values.

For plumes comprised of particulates with appreciable gravitational settling velocities, ISCLT accounts for the effects on ambient particulate concentrations of gravitational settling and dry deposition. ISCLT can also be used to calculate dry deposition.

ISCLT accepts stack, area, and volume source types. The volume source option is also used to simulate line sources. The steady-state Gaussian plume equation for a continuous source is used to calculate ground-level concentrations for stack and volume sources. The area source equation in ISCLT is based on the equation for a continuous and finite cross wind line source. The model can evaluate the effects of the aerodynamic wakes and eddies formed by buildings and other structures on plume dispersion.

A wind-profile exponent law is used to adjust the observed mean wind speed from the measurement height to the emission height for the plume rise and concentration calculations. ISCLT has the option of a rural mode or one of three urban modes. Time dependent exponential decay of the

pollutant can be included in the model. Options available to you include plume rise as a function of distance, stack-tip downwash, and buoyancy-induced dispersion. A regulatory default is available to allow you to direct the program to use features generally recommended by EPA for regulatory applications.

For more information of the proper use of the ISCLT model, you are urged to consult *Industrial Source Complex (ISC) Dispersion Model - User's Guide*, 2nd Edition, (EPA-450/4-86-005a). The NTIS order number for this document is PB86-234259/LP. The user is also referred to the Office of Air Quality Planning and Standards (OAQPS) document *Guideline on Air Quality Models* (NTIS No. PB88-150958) for recommendations on air quality modeling techniques.

3.1.1.2 TOXBOX

General Sciences Corporation developed TOXBOX for the EPA Office of Toxic Substances based on the work of Hanna (1972, 1977, 1980). The model is applicable to regions containing many diffuse emissions sources within their boundaries, such as an urban area source. TOXBOX is a steady-state box model, and assumes no upwind background concentrations. Dry deposition, precipitation scavenging, and chemical removal processes are accounted for in the continuity equation for material entering and leaving the box above the modeled regions.

TOXBOX computes a weighted average concentration based on stability frequency of occurrence. The stability frequencies and their corresponding average winds speeds are computed from running five-year annual average STability ARray (STAR) data compilations. Documentation for the TOXBOX model is in the *GAMS User's Guide* (GSCTR-32-88-017, 1988).

3.1.1.3 Input

GAMS models require the input of a significant number of data parameters. The GAMSATMOSET Procedure, available through the GAMS menu, prompts you for the information required to set up the desired modeling scenario. The data

requirements are summarized in Figure 3-1. Broken lines indicate optional transfers between prompt groups where the direction of transfer is based on your response to the prompts. Solid lines represent no option transfer between prompt groups.

The user interface for this procedure is prompt driven. The sequence of prompts is designed to transfer you from control prompts to chemistry prompts, and then through a sequence of prompt groups covering removal specifications, regulatory default options, meteorology, grid specifications, terrain elevations, source characterization, emission rates, and output specifications. A detailed description of the ISCLT and TOXBOX implementation within GAMS is given in the *GAMS User's Guide* (GSCTR-32-88-017, 1988).

3.1.1.4 Output

The output that can be generated within GAMS can be (1) model output which is the standard ISCLT and/or TOXBOX results written by the models, or (2) GEMS output which may take the form of GEMS graphs generated via GAMSGRAPH, GEMS reports generated via GAMSERE, and GAMS support files which are written into the user's directory and not into GEMS.

If you elect to produce the ISCLT or TOXBOX models output, the file names will consist of the run name followed by "ISC", followed by the three-digit site number for each of the ISCLT model output files; and the run name followed by "TOXBOX" for the TOXBOX model output file. ISCLT and TOXBOX model output files from the performance of GAMS are stored in files with the ".OUT" extension. These files are written to the user's directory and not into GEMS.

The content of ISCLT model output file is controlled by the user's selection of output options. You may control the type of input data that is to be written to the ISCLT output file by selecting one of the following options: none of the input data; only the control parameters, receptor points, and meteorological data; only the source input data; or all of the input data. Following the summary of the input data comes

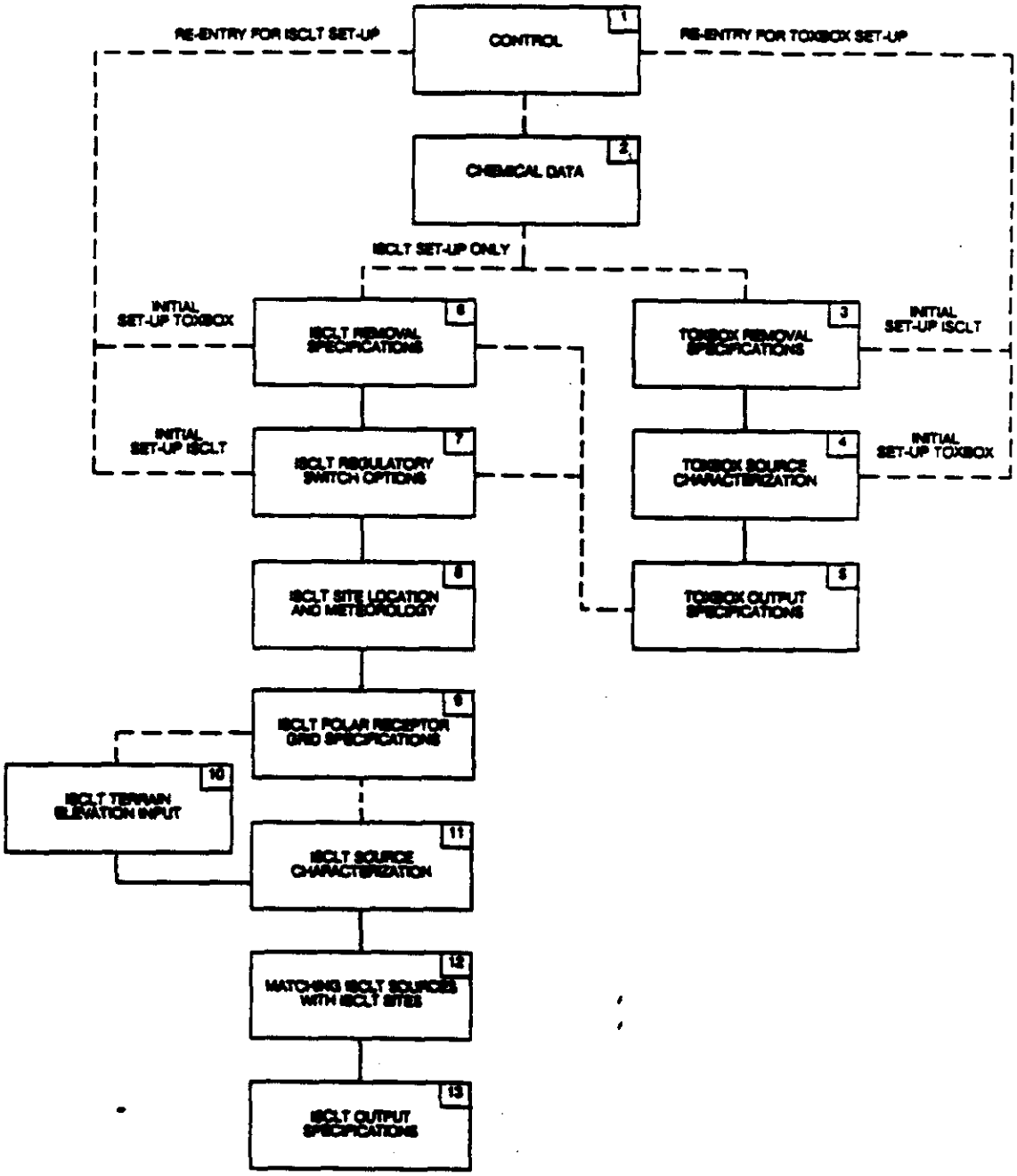

FIGURE 3-1. Sequence of GAMSIN Logical Prompt Groups

the tables of concentration estimates by direction and distance. These tables are listed by source and then there is a table of results for all sources. There is also an option to have the maximum 10 concentration estimates and their receptor locations printed.

The TOXBOX model output consists of a listing of the input parameter values and the concentration estimates for each UA and place for each modeled TOXBOX area source. There are roughly 4,000 UAs and places modeled by TOXBOX nationwide, so it is very strongly recommended not to produce the TOXBOX model output.

Two basic types of GAMS reports, Comptabs and Digests, can be generated from GAMSERE. The Comptabs are comprehensive tables of results that can be tabulated by a combination of source category and emission type, by source category, by emission type, and across all source categories and emission types. The exposure and dosage tables give cumulative population exposed by concentration level, and the risk tables give population exposed, order of magnitude risk levels, and excess cases of cancer.

The Digests contain total results compiled by source category. An Exposure Digest contains the maximum, minimum, and weighted average estimated concentrations; the total modeled population; and the total exposure by source category. The Risk Digest contains population at risk by order of magnitude risk levels and by source category. The Cases Digest contains the modeled population, the weighted average individual risk, the total number of estimated lifetime excess cancer cases, annual incidence, the maximum exposed individual risk, and the number of persons exposed at the maximum risk by source category.

GEMS reports created by GAMSERE are written to your directory, not to GEMS. GEMS graphs created by GAMSGRAPH are written to GEMS in your user dataset library. Consult the *GAMS User's Guide* (GSCTR-32-88-017, 1988) about the methodology used by GAMS to create the various file names and about the information contained in the GAMS output files.

3.1.2 INPUFF

INPUFF is a Gaussian plume, INtegrated PUFF model with a wide range of applications. The model is capable of addressing the accidental release of a substance over several minutes or of modeling the more typical continuous plume from a stack. The implied modeling scale is from tens of meters to tens of kilometers. Computations in INPUFF can be made for a single point source at up to 25 receptor locations. Up to 144 separate meteorological periods of the same length may be used to characterize the meteorology during the event. INPUFF is capable of simulating both moving point and stationary sources.

INPUFF utilizes three distinct dispersion algorithms. For short travel time dispersion (where the sampling time is long compared to the travel time), you have the option of using either the Pasquill-Gifford scheme or the on-site scheme developed by Irwin in 1983. The third dispersion algorithm is used for long travel time in which the growth of the puff is assumed proportional to the square root of travel time.

In 1984, W. B. Peterson, and some associates at EPA's Environmental Sciences Research Laboratory in the Office of Research and Development, wrote *INPUFF - A Single Source Gaussian Puff Dispersion Algorithm User's Guide* (NTIS No. PB85-137131). This publication contains more detailed instructions on the use of INPUFF. You may also refer to the *INPUFF User's Guide* (GSCTR8621, 1986).

INPUFF is located under the Air Models menu of GEMS. Two procedures are available under the INPUFF menu: INPUFFIN and INPUFFRUN. INPUFFIN creates the required INPUFF input data files using your responses to a series of menus and/or prompts. The INPUFFRUN procedure is used to execute the model.

3.1.2.1 Input

INPUFFIN is a sequence of menus or prompts that guide you through the set up of your modeling scenarios and create the INPUFF input file. Figure 3-2 illustrates this menu sequence

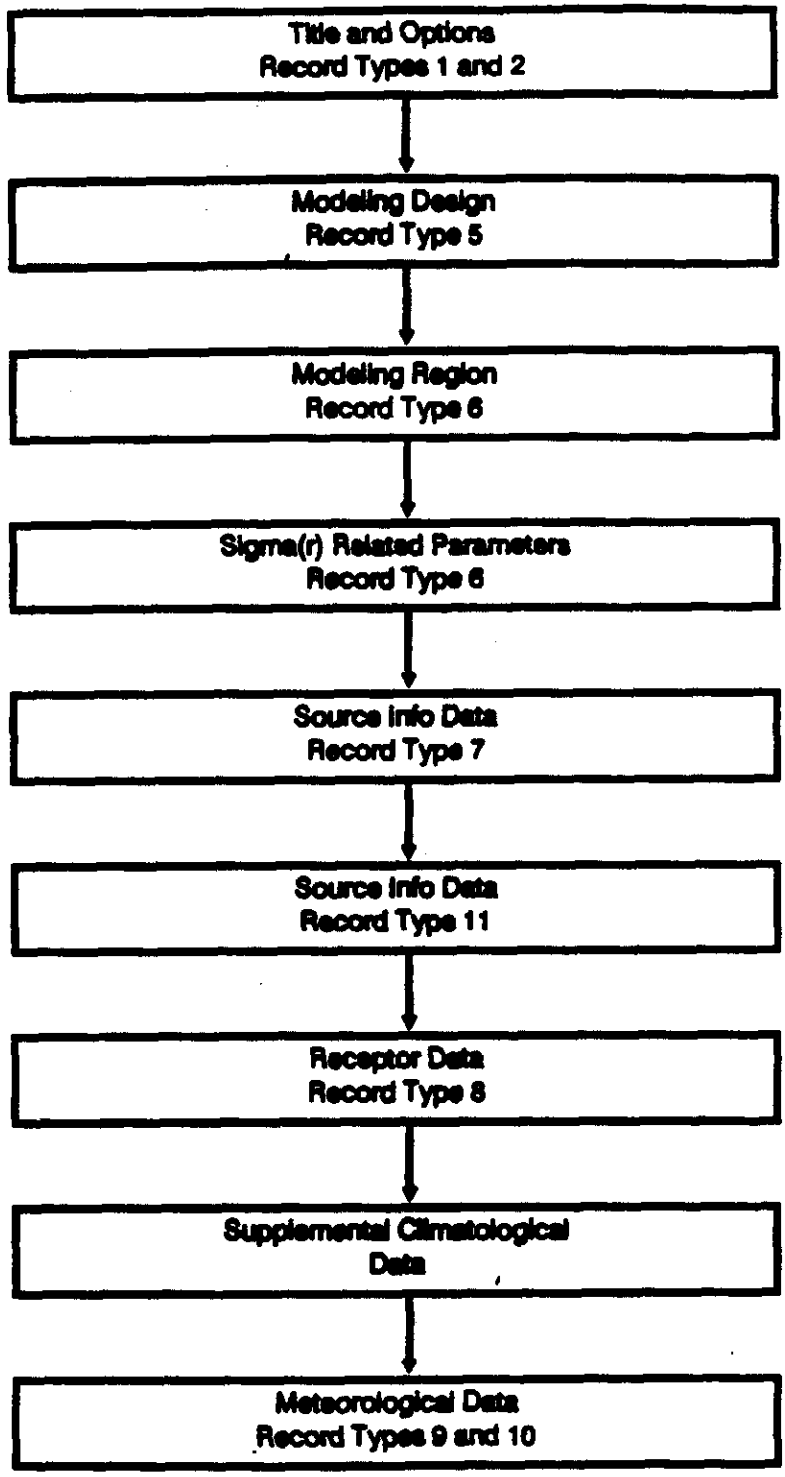

FIGURE 3-2. INPUFFIN Sequence

within INPUFFIN. It should be noted that supplemental climatalogical data can be accessed at the user's direction.

3.1.2.2 Output

The INPUFF model output is written to a GEMS report file. The report can be displayed at your terminal through the Output option in the File Management operation.

The output file begins with the report title, up to 60 characters, followed by a listing of the options and input parameters you provided. The report continues with a listing of the source data and the meteorological conditions used in the execution of the model. Next, the report gives you five parameters used by INPUFF to simulate the release:

- Simulation period
- Simulation time
- Puff release rate
- Minimum source-receptor distance
- Dispersion type

INPUFF reporting continues with two optional sections. The intermediate concentrations are optionally written at a user defined interval. The time period for which the averages are appropriate is printed in the first line of the intermediate concentration output. Information on each puff is optionally printed each time, in addition to average concentrations at each receptor. Finally, a table of average concentrations is produced.

3.1.3 PTPLU

PTPLU (from PoinT PLUme) is an improved model for estimating the location of the maximum short-term concentration from a single point source as a function of stability and wind speed. The algorithm is similar to PTMAX, which was first released in May 1973 (Turner and Busse, see references). Among the improvements in this version are

options for the estimation of gradual plume rise, stack downwash, and buoyancy-induced dispersion. Maximum concentrations and their corresponding downwind distances are calculated for two sets of wind speeds: winds assumed to be constant with height and winds assumed to increase with height. For the latter case, wind speed is extrapolated from anemometer height to stack top using a power-law wind profile. PTPLU is based on the point-source solutions of the Gaussian plume equations. Multiple reflections are considered until the vertical dispersion parameter is 1.6 times the mixing height; uniform mixing is assumed thereafter. No fumigation or chemical reactions are considered. In addition, GSC has implemented PTPLU so that concentration estimates can be obtained as a function of user-specified distances.

For more information on the use of the PTPLU model, we urge you to read *PTPLU - A Single Source Gaussian Dispersion Algorithm User's Guide*, (EPA-600/8-82-014, NTIS No. PB83-211235) and the *User's Guide for PTPLU in GEMS* (GSC-TR-32-88-019, 1988).

3.1.3.1 Input

PTPLU requires you to enter several data items before it begins. Data entry is in response to menus, and includes: use/don't-use options for gradual plume rise, stack downwash, and buoyancy-induced dispersion; ambient air temperature, mixing height, receptor's elevation above ground, anemometer height, wind-profile exponents, title, source strength, physical stack height, stack gas temperature, stack gas velocity, and stack diameter.

The PTPLU model has two modes of operation as implemented in GEMS: standard and distance. The standard mode produces a maximum concentration estimate for each stability and wind speed combination. The distance mode produces concentration estimates at up to twenty user-specified distances for each stability and wind speed combination. Either mode or both may be executed for each run of the model.

A control menu follows the initial sequence of menus. The control menu allows you to run the model with the current specifications or it allows you to go back to any of the menus and change the input specifications. After you are finished with a menu, you will go directly back to the control menu where you can run the model again or change values in another menu. The control menu also allows you to end the PTPLU modeling session.

3.1.3.2 Output

The results of the model are always sent to the terminal when the model runs in standard mode. After each standard modeling run of a session, you are asked if you want to keep the results. The results will be appended to a single GEMS report if you elect to keep them. Each modeling run conducted in distance mode within a PTPLU session is written to a unique GEMS dataset. The GEMS report and datasets may be reviewed using the OUTPUT procedure of GEMS. A menu will appear at the end of your GEMS PTPLU session requesting a report and/or root dataset name if either have been created. Each dataset name will consist of the root name you supply with a numeric suffix indicating the sequence number of the run within the PTPLU session.

The standard mode output includes: (1) a heading that indicates the program name and source, (2) a section that reiterates the input parameter values, (3) a section giving the calculated values for volumetric flow and the buoyancy flux parameter which are used for plume rise, and (4) the results of the model that include for each combination of wind speed and stability, the maximum concentration, the distance of the maximum, and the plume effective height at this distance. These results are given for the assumption of constant wind speed with height and for extrapolated wind speed with height.

Each dataset created by a PTPLU run in distance mode has four groups. The first group contains the variable names and the descriptions of all the variables in the dataset. Note that the units of the receptor distances are kilometers and the units of the concentrations are in micrograms per cubic meter. The second group contains the title. The third group gives the

values of wind speed by stability category for each of the 49 combinations. The fourth group contains the receptor distances and the concentration estimates at each receptor distance for each of the 49 wind speed by stability category combinations. The concentration estimates are written to the GEMS dataset are for the assumption of extrapolated wind speed with height.

3.1.4 BOXMOD

General Sciences Corporation developed BOXMOD for the EPA Office of Toxic Substances. This model is an interactive simple atmospheric area source model based on the work of S. R. Hanna, published in 1972 and 1977. You will find BOXMOD is most helpful when you are screening chemicals for potential health or environmental problems, especially in regions containing many diffuse emission sources, such as urban area sources. You can use BOXMOD for averaging periods of less than a day, but estimated concentrations are more nearly accurate using an averaging period greater than a few days up to annual averages.

You can use BOXMOD to calculate the atmospheric concentration for a single area where you input the location and width of the area. BOXMOD provides you with the option of a solution to the box model assuming either a uniform or Gaussian concentration distribution in the vertical, and you have the option of selecting STAR and auxiliary climatological data or supplying your own.

BOXMOD is a steady-state box model, assuming no upwind background concentration. Dry deposition, precipitation scavenging, and chemical removal processes are accounted for in the continuity equation for material entering and leaving the box above the modeled region. BOXMOD calculates a single pollutant concentration applicable to the entire region on a uniform area emission rate. The box is assumed square with a north-south, east-west geographic orientation.

BOXMOD is the stand-alone version of the TOXBOX urban area source model implemented in GAMS. Documentation for

this area source model can be found in the *GAMS User's Guide* (GSCTR-32-88-017, 1988).

3.1.4.1 Input

In order for BOXMOD to perform, you must supply information about location, weather conditions, and chemical behavior. Location information includes the latitude and longitude for the center of the area and the width and emission rate of the area. Weather conditions are specified by either the STAR station index number or wind speed, stability type, and precipitation rate and frequency under the Gaussian options, or wind speed, mixing depth, and precipitation rate and frequency under the uniform option. Chemical behavior is indicated using particle radius and density, or molecular weight if a gas, for estimating the dry deposition speed; and the time constant applicable to the chemical removal process.

3.1.4.2 Output

BOXMOD writes its output to a file in your VAX directory called FOR033.DAT (not a GEMS dataset) and to your screen. The output file contains a restatement of all of your input variables along with the estimated values for the scavenging coefficient, deposition speed and pollutant atmospheric concentration. After you have exited the BOXMOD routine, you can rename and/or use the output file as you wish.

3.2 Soil and Groundwater Models (SOIL)

GEMS provides two different models to estimate chemic concentrations in the unsaturated zone:

- SESOIL
- PRZM

as well as two models to estimate chemical concentratic the saturated zone (groundwater):

POOR COPY RECEIVED
9 2 1 2 3 1 9 1 2

- AT123D
- SWIP

SESOIL is a one-dimensional unsaturated soil zone model which simulates the vertical transport and transformation of a chemical. PRZM is an unsaturated zone transport model which simulates the vertical movement of pesticides in unsaturated soil within and below the plant root zone. AT123D is a saturated zone analytical transport model which simulates chemical movement in an aquifer in three dimensions. SWIP is a saturated zone three-dimensional numerical model which simulates contaminant movement in an aquifer, including the effect of well pumping.

3.2.1 Soil - SESOIL

SESOIL is a seasonal soil compartment model developed by M. Bonazountas and J. Wagner in 1984 for EPA's Office of Pesticides and Toxic Substances. This model is designed for long-term environmental pollutant fate simulations that can describe pollutant transport/transformation in the unsaturated soil zones.

General Sciences Corporation implemented SESOIL in GEMS and developed the input and output data management systems to support it. The input data management system uses menus to assist you in the preparation of SESOIL input data files. The output data management system allows you to store the SESOIL output data in GEMS. This in-GEMS storage allows you to utilize the file manipulation capabilities of GEMS for data analysis and graphic interpretation.

You will find SESOIL useful for a wide range of applications, including the screening and detailed review of chemicals. For example, it can be used to compare the effects of a large number of chemicals when released into a generic environment, it can provide necessary evaluative information in a site-screening management plan, or it can be used as a detailed, site-specific pollutant transport modeling tool. This model simulates a soil column, which you specify, and which extends from the ground surface to the saturated zone. The

soil column is divided into a number of compartments, each representing a homogeneous soil layer. Simulation of the transport/transformation processes within each compartment and between compartments are based on three cycles:

- Water cycle, which takes into account the hydrologic balance.
- Sediment cycle, which accounts for sediment washload (land erosion due to rainfall).
- Pollution cycle, which accounts for the transport and transformation processes directly affecting the fate of the pollutant.

SESOIL allows you to select from a number of operation options of modeling applications; each option associated with certain temporal and spatial resolution characteristics, and each having a different specific input requirement. The most commonly utilized level for OTS applications is Option M which performs calculations on a monthly basis and uses a four-layer soil column. Enhancements to SESOIL allow additional spatial resolution within the four soil layers by specifying up to ten optional sublayers for each.

For more information on the proper use of SESOIL, we suggest you read *SESOIL: A Seasonal Soil Compartment Model* by Marcos Bonazountas and Janet M. Wagner of Arthur D. Little, Inc (1984). This User's Guide is bound under Arthur D. Little's Number C- 85875 and was prepared under EPA contract No. 68-01-6271, for the Office of Toxic Substances. This document is available from NTIS (PB86-112406). An additional document, the *User's Guide to SESOIL Execution in GEMS* (GSC-TR8747, 1987), written by GSC for EPA's Office of Pesticides and Toxic Substances, provides a description and sample runs of the use of the SESOIL in GEMS.

3.2.1.1 Input

The SESOIL input files are created using the SEBUILD procedure. When you use SESOIL, you must provide climatologic, soil, and chemical specific data; pollutant application data file(s), containing information concerning the

size of the study area, depth to groundwater, ratio relationships among soil layers, and pollutant input from various sources; and a data file containing model execution information. The climate, soil, and chemical data may be accessed from various online databases available in GEMS.

GSC developed the input data management system (SEBUILD) to help you in the preparation of your input files. The flowchart in Figure 3-3 illustrates the execution of SESOIL. If you wish to execute SESOIL for an immediate run, using run ready input files, you may use the SERUN procedure.

3.2.1.2 Output

SESOIL produces a large quantity of output information, and GSC developed a file management utility for SESOIL that allows you to create GEMS datasets using the output from SESOIL.

SESOIL output includes:

- Resulting hydrologic relationships among precipitation, surface runoff, infiltration, evapotranspiration, soil moisture, and groundwater runoff (recharge).
- Temporal and spatial pollutant concentration distributions in soil-air, soil-moisture, and on soil particles.
- Leachate (pollutant mass) migration from the unsaturated soil zone to groundwater.

SESOIL writes its monthly and annual summary output to a report file which the user is asked to name. The user is prompted for an additional, optional GEMS dataset name, in which can be stored model results required for tabular or graphic output displays.

SESOIL also has the capability of creating an output file for AT123D, a saturated zone model (see Section 3.2.3), using computed monthly chemical mass quantities entering the

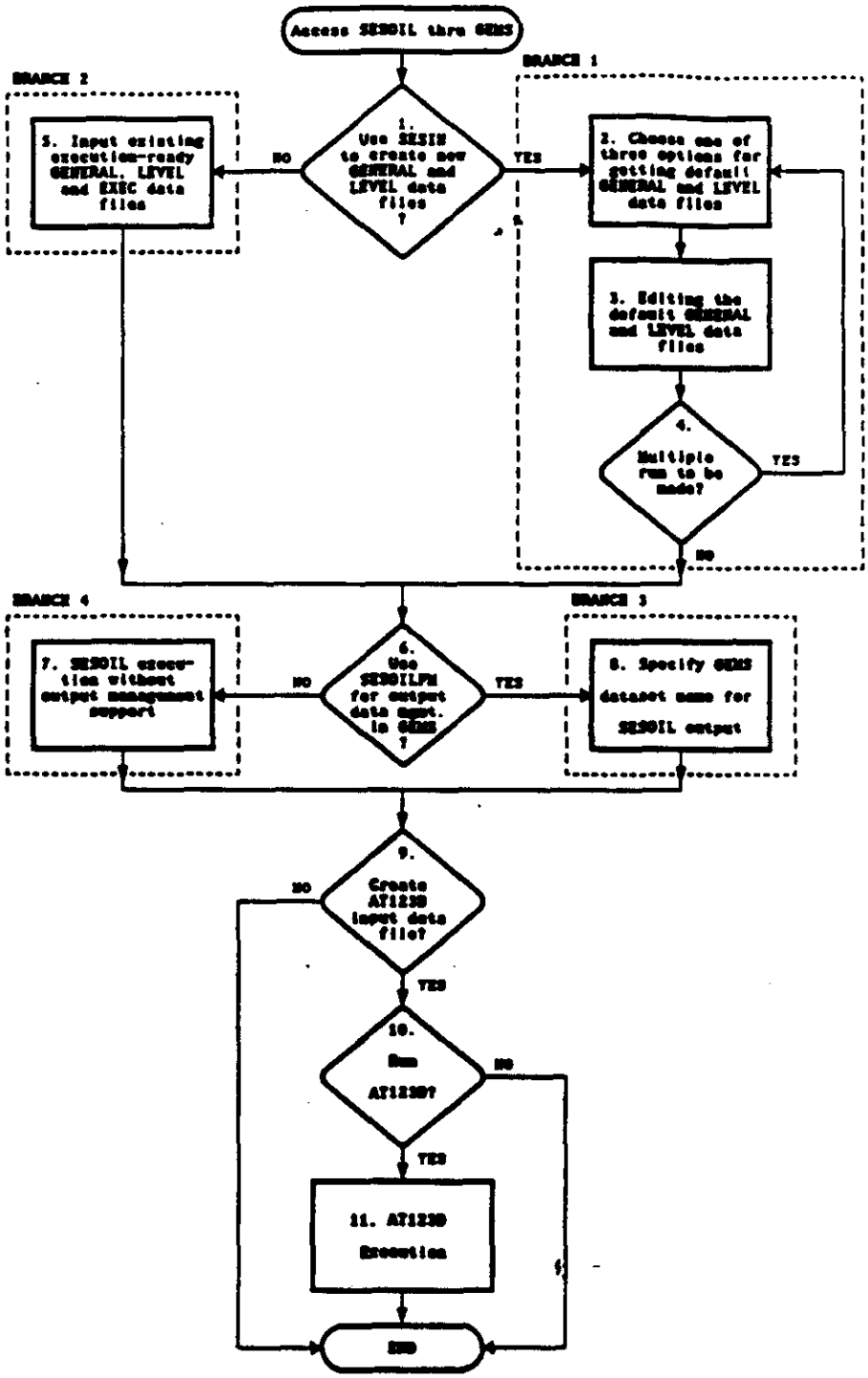

FIGURE 3-3. Flow Chart For SESOIL Execution With Data Management System Support

saturated zone and various other soil properties derived from SESOIL data. The SESATLINK procedure is used for this purpose.

3.2.2 Soil - PRZM

PRZM is a dynamic model which simulates the vertical movement of pesticides in the unsaturated soil, within and below the plant root zone, extending to the water table, using generally available input data that are reasonable in spatial and temporal requirements. PRZM consists of hydrology and chemical transport components that simulate runoff, erosion, plant uptake, leaching, decay, foliar washoff and volatilization (implicitly) of a pesticide. It was developed by EPA Environmental Research Lab at Athens, Georgia.

Time-varying transport, including advection and dispersion, are represented in the model. PRZM has two major components: hydrology and chemical transport. The hydrology component calculates runoff and erosion based on the Soil Conservation Service curve number technique and the universal soil loss equation. Evapotranspiration is estimated from pan evaporation data or by an empirical formula if input pan data are not provided. Evapotranspiration is divided among evaporation from crop interception, evaporation from soil, and transpiration from the crop. Water movement is simulated by the use of generalized soil terms including field capacity, wilting point, and saturation. Drainage from loose, porous and tighter compact soils is simulated. To produce soil water and solid phase concentrations, the chemical transport component calculates pesticide uptake by plants, surface runoff, erosion, decay, vertical movement, foliar loss, dispersion, and retardation. PRZM uses a finite difference numerical solution, using a backwards difference implicit scheme.

For more information on the proper operation of the PRZM model, we suggest you read *User's Manual for the Pesticide Root Zone Model (PRZM)*, Release 1, (EPA-600/3-84-109), written by Anderson Nichols, Inc. and the EPA Technology Development and Applications Branch, Environmental Research Laboratory, and published by the EPA

Environmental Research Laboratory at Athens, Georgia. This document is available from NTIS (No. PN85-158913/LT).

3.2.2.1 Input

You must accomplish three tasks to complete a simulation using PRZM.

- Develop the input stream for the model.
- Place the climatic data on a file in the format required for the model's use.
- Provide a time-series analysis of the output.

A special procedure, in FORTRAN, called ANPRZM provides you with a simple method for creating, checking, and updating input streams of data. To use this tool, you will still need to provide watershed and pesticide characteristics such as field capacity, wilting point, curve number, crop type, partition coefficient, and decay rate. Figure 3-4 illustrates the execution of PRZM using ANPRZM for setup.

3.2.2.2 Output

PRZM provides you with four output files. These are: daily, monthly, and yearly hydrological summaries; the pesticide mass balance; the pesticide concentrations over space and time; and time series output of selected parameters.

3.2.3 Groundwater - AT123D

The Oak Ridge National Laboratory developed this analytical transient one-, two-, and/or three-dimensional computer model. GSC implemented it in GEMS for EPA's Office of Toxic Substances. ATBUILD is a procedure developed by GSC to allow you to prepare the model input data file and ATEDIT is a procedure that allows you to edit existing input files. ATRUN executes the model using the input files that you enter or create. These three added procedures allow the integration of this model into the GEMS File Management System.

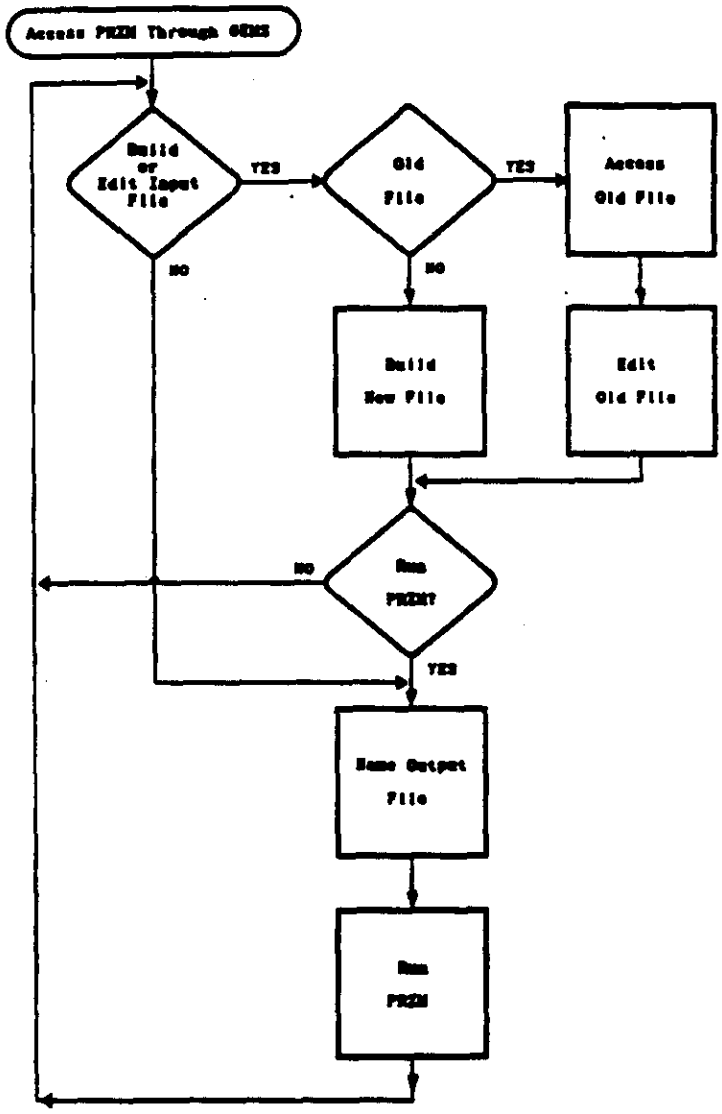

FIGURE 3-4. Flow Chart of ANPRZM Execution

AT123D is designed to provide an efficient and easy-to-use tool for the groundwater pollutant transport analysis. The model is capable of handling three different pollutant types: chemical, radioactive, and heat. It has been used from estimating the likely pollutant concentrations in the groundwater found downgradient from pollutant sources including domestic septic tanks and landfills. You may choose between eight types of pollution source configurations:

- Point source
- Line source parallel to the X-axis
- Line source parallel to the Y-axis
- Line source parallel to the Z-axis
- Area source perpendicular to the X-axis
- Area source perpendicular to the Y-axis
- Area source perpendicular to the Z-axis
- Volume source

The patterns of source releases acceptable to the model include instantaneous, continuous, and finite duration. Selection can also be made to define the aquifer dimension in terms of finite depth and finite width, finite depth and infinite width, infinite depth and finite width, or infinite depth and infinite width. Transport mechanisms available include advection, hydrodynamic dispersion, adsorption, and decay. The chemical leachate (mass entering groundwater) simulated in a SESOIL run can be used as the source loading rate in the AT123D model.

For more information on the proper use of the AT123D model, we suggest you consult *User's Guide to AT123D Execution* (GSC-TR8542, 1985) prepared by GSC for EPA's Office of Pesticides and Toxic Substances Exposure Evaluation Division. The original model documentation, *AT123D: Analytical Transient One-, Two-, and Three-Dimensional Simulation of Waste Transport in the Aquifer System*, by G.T.

Yeh (1981), is available from NTIS under order number ORNL-5602/LT.

3.2.3.1 Input

The AT123D input data management system takes you step-by-step through the process of defining the input file which must include the job title, model execution specifications, aquifer size and loading source size configurations, soil and waste properties, and the location coordinates where the solutions are desired. The input data file may be built either by editing default values or by accessing and editing an existing input file. An AT123D input file may be created from a SESOIL run using SESOIL mass to groundwater quantities as the source loading data for AT123D (see Section 3.2.1.2). Various other aquifer properties for AT123D are also derived from the SESOIL output. This input file may be accessed and modified as desired. Figure 3-5 illustrates input file creation.

If you are a PCGEMS user, you can build input files with your PCGEMS program and then submit your job to GEMS on the VAX for your AT123D modeling run. Consult your *PCGEMS Users Guide* for more information on this method of using AT123D.

3.2.3.2 Output

Model results are stored in a GEMS report file which is named by the user. Additional GEMS datasets may be stored, and accessed using the GEMS file management facility for your graphics display needs. The file contains:

- Summary of all input hydrologic, chemical, physical, soil, and execution specification parameters considered in the model simulation.

- Tabular presentation of the estimated pollutant concentration distribution in the groundwater at different time intervals and at different locations which you pre-specified along the X, Y, and Z axes.

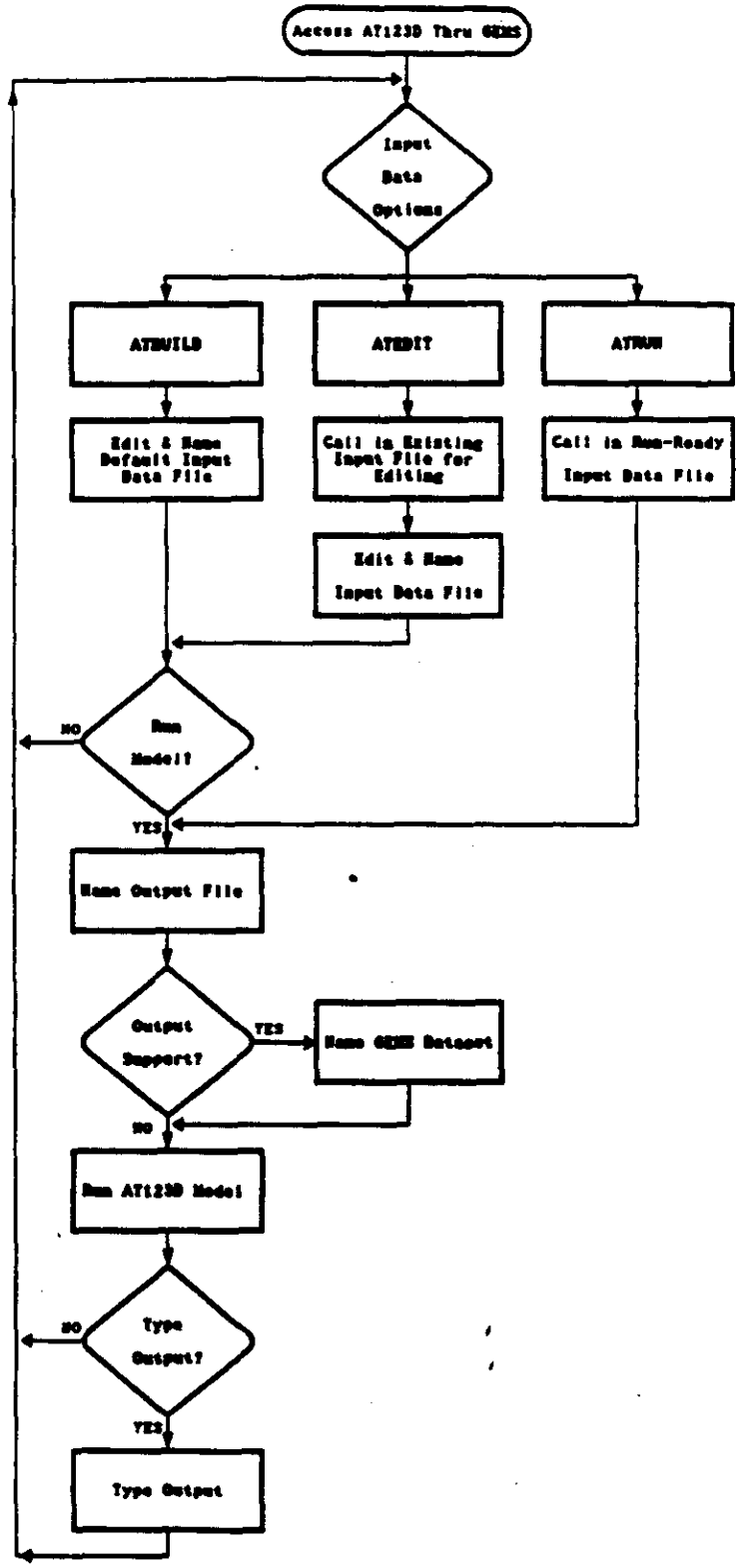

FIGURE 3-5. Flow Chart of AT123D Execution With Data Management System Support

The additional GEMS datasets store user-requested portio of this output for graphics.

3.2.4 Groundwater - SWIP

The Survey Waste Injection Program (SWIP) was origina developed by Intercomp, Inc. for the U.S. Geological Surv to investigate problems associated with the disposal of wast into deep wells. The model was later enhanced for mo general applications. The SWIP model is a three-dimension numerical computer model for simulating pollutant (chemic or thermal) transport/transformation in a groundwater syster with or without wells. GSC developed the input and output fi handling enhancements to the system and installed it in GEM for EPA.

SWIP incorporates the finite-difference approximation and ha options for several matrix solution techniques. It treats stead or transient flow in heterogeneous and/or isotropic media. also has options for treating flow in one-, two-, c three-dimensional Cartesian or cylindrical coordinates. SWI incorporates a wellbore model for calculation of pressures heat loss, and flow rates in the wells.

For additional information on the proper use of the SWIF model, we urge you to read the *User's Guide to SWIP Mode Execution Using Data Management Supporting Syster* (GSC-TR8531, 1985) written by GSC for EPA's Office o Pesticides and Toxic Substances, Exposure Evaluation Division. A detailed description of the original model design and capabilities is presented in Water Resources Investigations 76-71, available from NTIS under order number PB256903/AS. The revised documentation for the enhanced model is presented in Water Resources Investigations 79-96, and is available from NTIS under order number PB80-122542.

3.2.4.1 Input

SWIP operates from a single input file that contains all the data it needs:

- Program controls
- Fluid and aquifer properties
- Initial aquifer conditions
- Recurrent data
- Well plotting data

The program controls data consisting mainly of control parameters used to control the major options in the model such as selection of solution of equations, wellbore calculations, output listings, and dimensioning of arrays. The fluid and aquifer properties include aquifer and fluid properties, and grid system description data. The initial aquifer conditions describe the initial concentrations, temperatures, and fluid velocity in the aquifer. The recurrent data is read before each time step when you wish to change the well conditions, internal time step calculations, solution technique, aquifer recharge conditions, or output controls of rate simulation period. The recurrent data is repeated in a loop as many times as specified. The well-plotting data is needed for plotting pressures, temperatures, and concentrations in the wells; and is likewise repeated in a loop for each well specified. Figure 3-6 illustrates the SWIP input data file preparation.

3.2.4.2 Output

The output file contains the pressures, temperatures, and concentrations in the specified aquifer grid block centers at each specified time step. Total quantities of water, heat, and contaminant mass pumped by each well for each specified time step are also provided. Contour maps of pressure, temperature and/or concentration at any specified time steps can be obtained, if you need them, as can plots of any or all of three quantities in each well versus time. The contour maps and well plots are outputted in character mode.

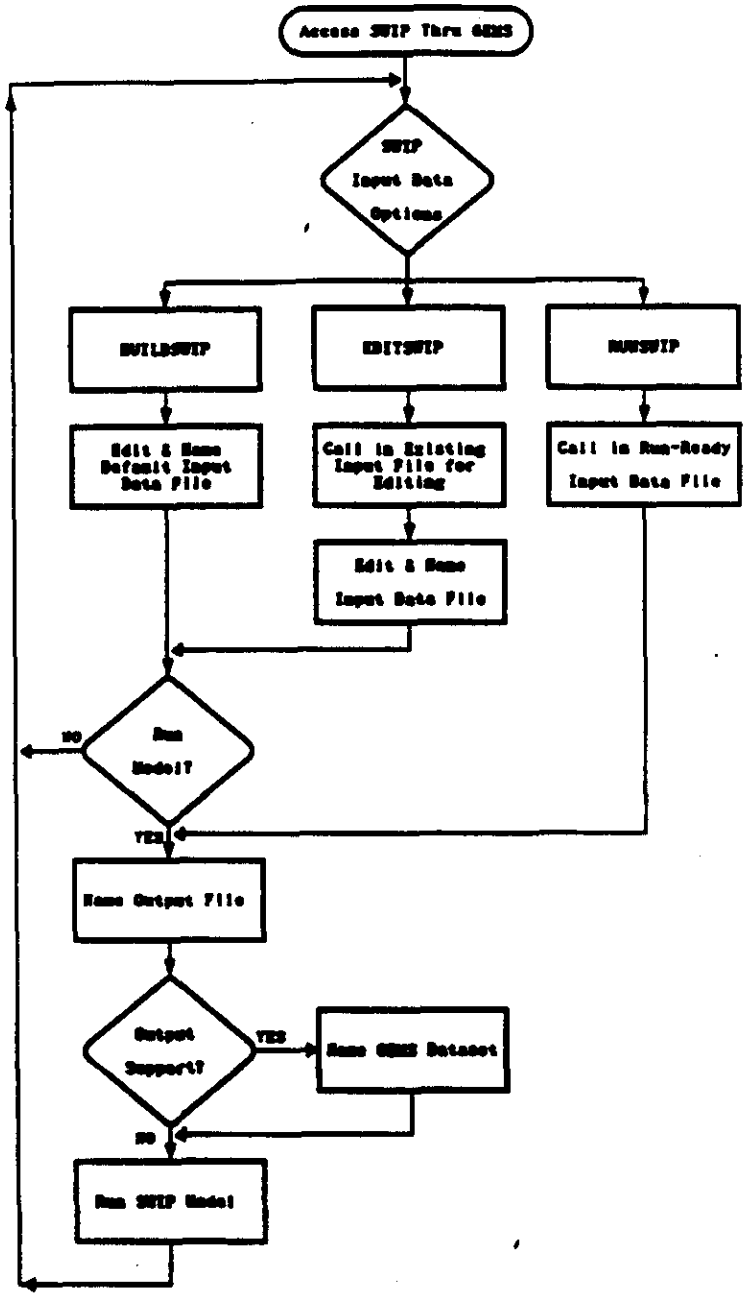

FIGURE 3-6. Flow Chart for SWIP Execution With Data Management System Support

3.3 Surface Water Models (WATER)

3.3.1 Surface Water - Exposure Analysis Modeling System II (EXAMS II)

EPA's Environmental Research Laboratory in Athens, Georgia developed EXAMS in 1980 for the rapid evaluation of the behavior of synthetic organic chemicals in surface water based on the relevant environmental characteristics of the ecosystem, and the properties of the chemicals. EXAMS simulates the steady-state distribution and concentration of the compound resulting from a specified pattern of release into either a lake or riverine ecosystem. EXAMS has been used for simulating steady-state chemical concentrations, at various locations in rivers, lakes and estuaries due at various locations in rivers, lakes, and estuaries due to upstream point source and non-point source discharges.

EXAMS II, the second version, was developed to enhance the modeling capabilities and options. This enhancement allows you to select different chemical release scenarios, including long-term, steady-state analysis, pulse analysis, or spike analysis, which performs simulation based on your specification of time-varying conditions and time factors. At the completion of a simulation run, EXAMS II provides options for plotting and/or listing of a variety of parameter combinations.

EXAMS II is a deterministic model based on a set of process equations. These equations account for the interactions between the chemistry of a compound and the environmental forces that shape the compound behavior in the ecosystem. The loading, transport, and transformations of a compound are balanced in the model in accordance with a mass conservation law. The ecosystem is represented by a set of N-compartments; each depicting a unique, constant, uniform entity of the ecosystem.

If you need to know more about EXAMS-II, we suggest you read *User's Guide to EXAMS II Execution in GEMS*

(GSC-TR8710, 1987), written by GSC for EPA's Office of Pesticides and Toxic Substances, Exposure Evaluation Division. A complete description of the original model can be found in *Exposure Analysis Modeling System (EXAMS): User Manual and System Documentation*, EPA-600/3-82-023, which is available from NTIS under order number PB82-258096.

3.3.1.1 Input

You have to create three input files before running EXAMS II. These files can be built using the EXAMS menu templates:

- Environmental data file
- Chemical data file
- Loading data file

The environmental data file contains information describing the environment in terms of a climactic, biological, hydrologic, and sedimentary characteristics. Each of the variables required by this data file has a direct bearing on the compound behavior. Monthly values may be used for environmental parameters or they may be constrained to single time-invariant values within a given compartment of the ecosystem. These descriptive variables are regarded as representing average or typical values of the environment. The chemical data file contains information related to the various chemical characteristics of the compound of interest. Examples of those data requirements include compound solubility, molecular weight, partition coefficients, hydrolysis rate constants, and biodegradation rate. The loading data file provides information concerning the compound loading to the ecosystem, as well as to the system background loadings in stream flow, rainfall, interflow, nonpoint source flow, and drift. The loading data file is stored as a file of EXAMS commands which may be invoked to create the loading data in the model.

EXAMS II has an interactive interface to the Canonical Environments Data Base (CEDB), providing direct access to environmental data describing the ecosystem for a variety of streams and lakes. You may access the input data for data

storing, updating, and retrieval via simple commands. Figure 3-7 is a system flow chart for EXAMS II in GEMS.

The CEDB is composed of two separate sections, one for lake and reservoir data, and the other for riverine data. Both of these sections meet the EXAMS environment data requirements. These two types of water bodies are separated because of the differences in their ecosystem configurations. Different variables are used to define these different types of waterbodies. The CEDB contains its own data base management system along with its own set of commands for user interface. The CEDB was, however, implemented in GEMS in such a way that you may access the desired data using GEMS commands alone.

For more information on the CEDB, you should consult Section 3.1 of the *User's Guide to EXAMS II Execution under GEMS* (GSC-TR8710, 1987).

3.3.1.2 Output

EXAMS II can provide a wide range of output products, including up to 20 different tables summarizing the following input data used in the estimations: a time trace of the average concentrations during the simulation period; distributions among aqueous chemical species at various times; locations of the minimum and maximum concentrations at various times; the fate of the compound in the form of a percent of loading, half-life, and mass flux expressed for each of the transport and transformation processes considered; a post load decay kinetics analysis, and the ultimate chemical concentrations and distribution in the ecosystem and the time required for the system self-purification after cessation of the chemical loading. A more detailed discussion of these output tables is offered in *The Training Guide and Reference Manual for EXAMS-II* by L.A. Burns and D.M. Cline of the Office of Research and Development, EPA (1985), available under NTIS PB82-258096.

GEMS datasets, which contain selected output concentrations in the ecosystem, may be created from an EXAMS run. These

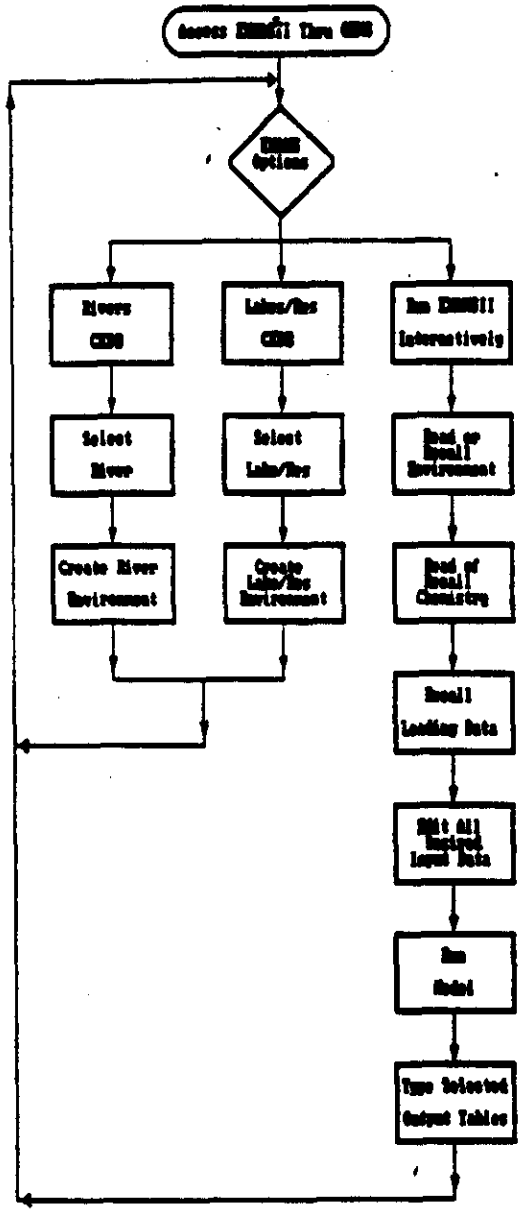

FIGURE 3-7. Flow Chart Showing EXAMS Execution

datasets may be used with the GEMS graphical and statistical facilities to produce plots, barcharts, and statistical analyses of the EXAMS output results.

3.4 Multi-Media Models (MULTI)

GEMS provides four models that can handle estimation of pollution in more than one media. These multi-media models include:

- ENPART
- TOX-SCREEN
- MICROBE-SCREEN
- UTM-TOX

ENPART is an environmental partitioning model which simulates the partitioning of a chemical in the environment between, air, water and soil. TOX-SCREEN is a multi-media screening level model which simulates the transport and interaction of a chemical in the air, in the unsaturated soil zone, and in a waterbody. MICROBE-SCREEN is similar to TOX-SCREEN, except that the "pollutant" modeled is a micro-organism, rather than a chemical. UTM-TOX is a multi-media simulation model which simulates the transport and interaction of a chemical in the same three media as TOX-SCREEN, but using more detailed site-specific data.

3.4.1 Environmental Partitioning Model (ENPART)

GSC worked with EPA's Office of Toxic Substances to develop ENPART, an environmental partitioning model designed to perform a first-level screening analysis from a minimum set of chemical- specific input data. This analysis is designed to help you identify the most important environmental pathways, the estimation of the substances' environmental persistence, and their bioconcentration potential.

ENPART is an adaptation of the fugacity approach to equilibrium partitioning as described by D. McKay in 1979. In this approach, the equilibrium concentration ratios between environmental compartments are calculated from simple physical/chemical data, such as molecular weight, solubility, and vapor pressure. In ENPART, this approach was modified by the addition of environmental parameters, such as compartment size, intercompartmental transport, and degradation, to yield a second level of concentration and mass partitioning. Both concentration and mass ratios are presented.

The environment that ENPART models consist of three compartments: air, water and soil. The water compartment is subdivided into four components: water, suspended sediment, biota, and bottom sediment. For the last three components, only equilibrium processes are considered. ENPART can simulate either equilibrium or dynamic partitioning. In equilibrium partitioning, only equilibrium processes defining the time-independent "steady state" environmental conditions are allowed to determine the estimated final distribution of the chemical. In dynamic partitioning, as used in ENPART, the chemical distribution is controlled by compartment-related transformation rates and intercompartmental chemical transport.

For more information on the proper use of the ENPART model, we suggest you read A *User's Guide to Environmental Partitioning Model* (GSC-TR8518, 1985), written by GSC for EPA's Office of Pesticides and Toxic Substances, Exposure Evaluation Division.

3.4.1.1 Input

A sequence of prompts elicits the information needed by ENPART. Many of the ENPART input parameters can be estimated within the model or are assigned default values. A more detailed description of the ENPART input procedures is contained in the model documentation published in 1982 by Wood, et al., and in 1985 by Jim Pilotte.

If you are a PCGEMS user, you can build ENPART input files with PCGEMS and submit the modeling job to the GEMS VAX system. Consult your *PCGEMS User's Guide* for more information.

3.4.1.2 Output

ENPART is designed as an interactive tool, and the output is written to your terminal screen while you watch. The output is also written to a file named "SUMOUT.TRM" in your default directory. You can rename or manipulate this output file to meet your needs. The file, and the displayed output, consist of intercompartmental concentration and mass ratios, concentration and mass partitioning factors, and a ranking analysis for both equilibrium and dynamic partitioning. Dynamic partitioning provides an estimate of the chemical persistence.

3.4.2 TOX-SCREEN

D. M. Hetrick and L. M. McDowell-Boyer developed TOX-SCREEN for EPA's Office of Pesticides and Toxic Substances in 1982. It is a multimedia screening-level model which assesses the potential fate of toxic chemicals released to the air, surface water, or soil. This model is more complex than ENPART since it incorporates equations developed for media-specific models and allows the user to describe the compartments directly. TOXSCREEN, therefore, requires more input than ENPART does.

Four types of surface water bodies may be considered: lakes, rivers, estuaries, and oceans. TOX-SCREEN is simple in nature and is intended to be used as a screening device to identify chemicals that are unlikely to pose problems even under conservative assumptions. GSC implemented TOX-SCREEN in GEMS after providing enhancements to manage the input data file.

The multi-media nature of TOX-SCREEN allows the model to simulate physical/chemical processes which drive transport of chemicals across air-water, air-soil, and soil-water interfaces. Such media interactions are handled with the use of deposition

velocities, transfer rate coefficients, and mass loading parameters. Monthly pollutant concentration in air, surface water, and soil reflect both direct input to any or all of the media from a specified source and subsequent interaction via processes such as volatilization, atmospheric deposition, and surface runoff. You must select the types of water bodies, if any, to be considered in any given simulation. You must also specify whether the pollutant is directly released into the air or water, or directly applied to soil.

For atmospheric dispersion simulation, a modified Gaussian plume equation was adapted for use in estimating downwind concentrations of a particular chemical emitted from a point source. A simple urban diffusion model was adopted for use in estimating ground-level pollutant concentrations over area sources.

For aquatic dispersion, an equation similar to that used in the EXAMS model was used to estimate the monthly pollutant mass in rivers and lakes. A one-dimensional steady-state model was used for simulating dispersion of pollutants in estuaries and a steady-state Gaussian-type linear diffusion model used in oceans.

TOX-SCREEN uses an adaption of the three-layer SESOIL model to estimate pollutant concentrations in the soil following introduction via direct application and/or interaction with other media.

For more information on the proper use of the TOX-SCREEN model, we suggest reading the *User's Guide to TOX-SCREEN Execution in GEMS* (GSC-TR8751, 1987). You are also urged to consult the *User's Manual for TOX-SCREEN: A MultiMedia Screening-Level Program for Assessing the Potential of Chemicals Released to the Environment* (EPA-560/5-83-024) written by Oak Ridge National Laboratory for EPA's Office of Toxic Substances. This document is available under NTIS order number PB84-213750/LT.

3.4.2.1 Input

TOX-SCREEN needs seven different input files to function:

- SESOIL EXEC data file
- SESOIL GENERAL data file
- SESOIL LEVEL3 data file
- Model Flags data file
- Air Parameters data file
- Water Parameters data file
- Bioaccumulation Parameters data file

GSC has developed a front-end for this model that facilitates the preparation of these requisite files. It is accessible from the TOXSCREEN menu, and it provides step-by-step assistance in the formulation, editing and naming of TOX-SCREEN input data files.

3.4.2.2 Output

TOX-SCREEN generates up to seven different output products for the simulation run. These files vary depending on just what you ask for. They will include any or all of the following:

- Results generated by SESOIL
- Summary of TOX-SCREEN input data
- Results generated for the water body (whether it is a lake, river, estuary, or ocean)
- Results generated from the food chain bioaccumulation calculations

The results include monthly chemical concentrations and mass distributions in the soil profile, monthly chemical concentrations at various locations in the water body, chemical concentrations in the air, and monthly inter-media interaction quantity (i.e., deposition on soil, volatilization from soil to air, groundwater recharge, etc.).

3.4.3 MICROBE-SCREEN

MICROBE-SCREEN is a screening level multi-media model developed to assess the potential fate of micro-organisms released to air, surface water, or soil. The model deals specifically with passively-dispersed micro-organisms, such as bacteria. MICROBE-SCREEN was adapted from TOX-SCREEN, a chemical fate model, by modifying TOX-SCREEN to account for processes important to modeling micro-organisms, such as growth and decay, aerosolization efficiency and attachment/detachment to the soil. (TOXSCREEN processes relating to chemical transformations were removed from this model.) MICROBE-SCREEN considers four types of surface water bodies: lakes, rivers, estuaries, and oceans. The model is simplified in nature and is intended to be used as a screening device to estimate transport and densities of micro-organisms in various environments.

GSC has implemented a data management system for management of both input and output data in support of MICROBE-SCREEN. More specific information on the GEMS implementation of this model may be found in the *User's Guide to MICROBE-SCREEN Execution in GEMS* (GSC-TR8753, 1985).

3.4.3.1 Input

MICROBE-SCREEN input is solicited from you by a series of interactive prompts and parameter menus. This interaction builds the requisite input files for the model. Six different files are required: GENERAL, LEVEL3, EXEC, FLAGS, AIR, and WATER.

The GENERAL file consists of monthly climatic, soil and microbial properties for the soil profile. The LEVEL3 file consists of the soil-layer and application loading parameters for a three-zone soil column. The EXEC file includes parameters which control the reading and execution of the GENERAL and LEVEL3 files. The FLAGS file consists of 11 model flags which control the water body and air data options. The AIR file holds the parameters describing the source

release to the atmosphere and factors affecting dispersion and settling. The WATER file contains the parameters which describe the water body environments, pollutant release rates to the water bodies, and properties of the microbe in the water body environment.

3.4.3.2 Output

Up to seven output files may be created by MICROBE-SCREEN, depending on a particular application. For all simulations, the SESOIL output file is produced. This file contains the output from the SESOIL submodel, including the monthly hydrologic balance and pollutant concentrations in the soil profile. Another file produced with each run is the "echo" of all input data from FLAGS, AIR, and WATER input files. This file includes error messages for any incorrect data in these input files. If water bodies are considered in the simulation, a separate output file is produced for each water body considered. These files include microbe concentrations in the air, water bodies and soil compartments, deposition rates on the water and soil surfaces; and surface run-off and surface recharge rates. If no water bodies are considered, then a "soil-air" output file is produced including air and soil concentrations.

3.4.4 UTM-TOX

The Oak Ridge National Laboratory developed the Unified Transport Model for Toxic Materials (UTM-TOX) for the EPA in 1983. It is a multimedia model designed for predicting the dispersion of pollutants through air, soil, and water. This model is more complex than either ENPART or TOXSCREEN. It links together several media-specific models. The user is therefore required to provide a large amount of input information.

The model calculates the atmospheric dispersion of up to 20 chemicals from a maximum of 10-point, 10-line, and 10-area sources. UTM-TOX also calculates deposition of one chemical at a time in both wet and dry form on foliage or the surface of the earth followed via the transport through runoff and erosion,

percolation through the soil to a stream channel, and transport in the stream channel to the outfall of a watershed.

During each of these phases of transport, many chemical processes can be represented in the model, including exchange with the soil, solubilization, degradation, volatilization, photolysis, and hydrolysis. Each of these processes is represented in the model by the incorporation of three previously developed models: ATM, WHTM, and SEDMNT. These models treat atmospheric, hydrologic, and sediment transport of pollutants, respectively. With UTM-TOX you can track a pollutant through the ecosystem, tabulate budgets for the partitioning of the contaminant, calculate the concentration of the pollutant in many compartments of the ecosystem, and reliably assess the impact of the pollutant.

ATM calculates the dispersion of airborne pollutants. It is a Gaussian plume model that determines the spreading of material emitted from a point source, a line source, or an area source. This determination is from calculations of the sigma values, and the vertical and horizontal widths of the plume. These sigma values are determined empirically and are representative of plumes of less than 50 km in length.

When the program is used in the annual or monthly mode, 16 sectors are used to describe the direction from which the wind blows. The concentration of the pollutant is assumed to be uniformly aximuthally distributed over the whole sector because of the meandering of the plume.

WHTM calculates the transport of pollutants through portions of the watershed. In the hydrologic cycle, the continuous processes of transpiration and evaporation driven by precipitation lead to interflow beneath the surface of soil, recharge and discharge of an aquifer, and establishment of flow in the stream channel. The model provides a parametric summary of these that is completely deterministic and very useful for purposes of simulating runoff and transport. The modeled cycle also includes uptake by vegetation, erosion, solubilization of absorbed material from the litter, transport from the litter, and infiltration into an unsaturated region of the soil where the root zone occurs.

SEDMNT models the aquatic transport of pollutants by sediment. This mode of transport is important because of sorption/desorption of a chemical on sediment can determine its fate. SEDMNT considers each reach of a stream channel to have a suspended-load layer that travels with the stream; a bed layer that is deposited at the bottom of the channel and is subject to resuspension; and a resident-bed layer that is more permanent. The fluxes among these various layers are calculated with (1) Stokes's Law fall velocities, (2) the shear stresses within each layer, (3) the particle sizes of the sediment, and (4) the load- carrying capability of the stream channel. The thickness of the different layers can change during the course of a simulation in response to the sediment deposit and/or resuspension in the progress of stream flow transport.

If you would like to know more about the proper use of the UTM-TOX model, read *Characterization of Data Base Requirements for Implementation of UTM-TOX Under GEMS: Parameter Sensitivity Study*. The original model documentation, *A User's Manual for UTM-TOX, the Unified Transport Model*, is available from NTIS (Order No. ORNL-6064/LT).

3.4.4.1 Input

UTM-TOX reads ten input files to set the parameters for the simulation run. These files provide monthly wind roses, hourly precipitation, solar radiation, daily maximum and minimum temperatures, soil characteristics, topographic information, surface water characteristics, sediment characteristics, and the chemical properties and transformation rates associated with the chemical. You will be prompted to provide identification for the ten input data files.

3.4.4.2 Output

UTM-TOX provides plots and tables as output. These plots and tables summarize all of your input parameters as well as the average monthly and annual chemical concentrations in wind sectors, in saturated and unsaturated soil layers, in runoff,

and out of each reach. The output is produced in three files in your logged directory:

- UTMTOX1.OUT
- UTMTOX2.OUT
- UTMTOX3.OUT

Chapter 4 - Geodata Handling

The Geodata Handling (GH) operation in GEMS provides access to geographically-linked information in GEMS datasets and other related files. For example, using one of the procedures, you may retrieve the population around a specific site. These types of retrievals are very useful in gathering in-depth information on specific areas of the county from the many databases available in GEMS.

To access this operation, type GH at the main GEMS navigational menu. The menu shown below appears.

```
MENU: Geodata Handling

1. Geographic Data Lookup and Display          (GEOLIST)
2. Mapping                                     (GEOMAP)


Enter an option number or a procedure name (in parenthesis) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 4-1. Geodata Handling Menu

The Geodata Handling operation has two major components: GEOLIST and GEOMAP. GEOLIST allows you to obtain specific information from selected datasets and the GEOMAP operation is used for mapping data. The procedures available from these operations will be described below.

When using the procedures available under the Geodata Handling operation of GEMS, you will often be asked to enter the geographic coordinates for the location that you want. You

may use a variety of geographic location methods to specify the point that you want:

ZIPCODE
This geographic location method tends to be the easiest of the geocodes to locate and use. The zip code is translated into the corresponding latitude/longitude coordinates for the post office servicing that zip code. The drawback to this is that this may result in latitude/longitude coordinates for a location, which while close to the location that you want, might not have the same geographic profile. This method of site specification should only be used when better data is unavailable.

UTM
The Universal Transverse Mercator coordinate system can be used for most GEMS procedures where geographic coordinates are needed.

LATITUDE/LONGITUDE
Two methods of specifying the latitude and longitude of a site are provided within GEMS. You can specify a coordinate using decimal degrees or degrees, minutes and seconds.

For a number of the following procedures, you may be asked to enter a Federal Information Processing Standard (FIPS) code for a state or county. This coordinate system allows you to identify a state or county in the United States. The state FIPS code is a two-digit code while the county FIPS is a three-digit code which when appended to the state code uniquely identifies a county. The FIPS codes for states are provided in the back of this user's guide in the section entitled Tables. Please note that in the sections below, geographic coordinates will be referred to as geocodes.

4.1 Geographic Data Lookup and Display (GEOLIST)

The Geographic Data Lookup and Display operation gives you the ability to scan a number of geographically organized

datasets and to manipulate data which has geocodes as part of the information in the dataset.

Each of the following procedures allows you the option of providing input necessary for the procedure either directly from your terminal or from a GEMS dataset. Use the terminal option if you are scanning a small number of sites. The dataset option should be used when you are retrieving data on a large number of sites, and when the necessary information to locate those sites has been previously stored in a GEMS dataset. When using the GEMS dataset input option, you have the option of saving the output in either an output report or a dataset. Since no output is directed to the terminal if you use the dataset input option, you must save the output.

You have two options when saving the output: you may save the output in a dataset and in a report. The difference is that when you save it as a dataset, the output is saved in a regular GEMS dataset format with variables and it may be used as input for other procedures. If you save the output as a report, the output is saved in a simple text format and is useful if you wish to print the output. The procedures above also have the option of running interactively or in batch. If you are using the dataset input option, then directing the output to a dataset or report (or both) is necessary. As there is no interactive display of data when using the dataset input mode, we recommend that you use batch to run the procedure if you use dataset input mode. This will free up your terminal for other work while GEMS processes your retrieval request.

The procedures below are discussed in detail in the following sections. In the following examples, zip code entry is used in the example menus. If you select the UTM option, the zip code entry in the following menus will be replaced by the entry of three variables: ZONE, NORTHING, and EASTING. Similarly, if you select the latitude/longitude entry, the zip code variable will be replaced by the variables asking for LATITUDE and LONGITUDE. Remember that zipcode is the least accurate method of identifying site locations.

```
MENU: NEW Geodata Handling Mapping Procedures

1. Site level retrieval of data                (SITERET)
2. Access Census data                           (CENSUS)
3. Determine County Coverage                    (COVERAGE)
4. Geographic Data Management                   (GEODM)
5. HUCODE/SOIL locator                          (HUCODE)
6. Convert to Lat/Long                          (LATLON)
7. Lookup/Examine Star Station Data             (STAR)
8. Find US cities                               (USCITY)

Enter an option number or a procedure name (in parenthesis) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 4-2. Main GEOLIST Menu

4.1.1 Site Level Retrieval of Data (SITERET)

The SITERET procedure is useful for retrieving all of the data available within the Geographic Data Handling operation on a site. This procedure produces a report containing the information from a number of separate retrievals. This procedure is equivalent to aggregating all the data that is output by each of the following GH procedures separately: CENSUS, COVERAGE, HUCODE, LATLON, STAR, and USCITY. (Each of these procedures is documented in a following section of this chapter.)

This procedure can accept input from either your terminal or from a GEMS dataset. The procedure can also be used to create an output GEMS dataset or report in your user dataset library. When you select the SITERET procedure, the parameter entry menu shown below appears. It is followed by a discussion of each parameter.

```
MENU: INPUT/OUTPUT SPECIFICATION

ref parmname   parameter description          value  index
1. METHOD      Input method specification     DIRECT
2. DATASET     Name of the output dataset     BROWSE
3. DTSTAG      Tag field of the output dataset *
4. REPORT      Name of the output report      BROWSE
5. RPRTAG      Tag field of the output report *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command:HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-3. Input/Output Specification (SITERET)

1. METHOD: The input method defines the source of the input to the procedure. The default is DIRECT which runs the procedure interactively with the input coming from your terminal. The other option is DATASET which indicates that the input is to come from a GEMS dataset. Note that if you select the dataset input, you cannot select BROWSE for output. You must specify a dataset or report name, which directs the output to a GEMS dataset or report.

2. DATASET: The name of the output dataset. If the default of BROWSE is selected, then results appear on your terminal and no output dataset is created. If you want an output dataset, you must enter the name of the dataset you would like the output stored under in your user dataset library.

3. DSTAG: The tag field of the output dataset. The tag field is a 60-character field which can be used to determine the contents of a stored dataset.

4. REPORT: The name of the output report. If the default of BROWSE is selected, ther results appear on your terminal and nc output report is created. If you want ar output report, then you must change the name of the report to the name you woulc like the output stored under in your use dataset library.

5. RPTAG: The tag field of the output report. The tag field is a 60 character field which can be used to determine the contents of a stored report.

Remember that you do not have to create both an outpu dataset and an output report. To create only an output datase or an output report, enter a name for the one that you wan and leave the other one at the default value of BROWSE. I you only wish to interactively view the data, you do not have to make any changes in the parameter entry menu showr above. Once you have made your selections, enter NEXT to move onto the next menu where you will specify the coordinate system to use to specify the center of the area of interest.

```
MENU: Select Method to Locate the Area of Interest

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude (decimal degrees)
4. Latitude/Longitude (degrees, minutes, seconds)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS
```

FIGURE 4-4. Geographic Coordinate Selection Menu

Select the method you will be using to specify the coordinates for your area of interest. The menus that appear for the various methods are similar except for the geographic code specification. Therefore, only the menu that appears for the zip code will be shown. For this example, the input method is DIRECT.

```
MENU: Site retrieval by ZIPCODE coordinate

ref parmname    parameter description                value  index
1.  ZIPCODE     Zip code of the study site
2.  SITENAME    Name of the study site
3.  STATE       State identifier
4.  DATA        Name of the Census database
5.  STANRING    Use standard ring distances          YES
6.  SECTORS     Number of sectors                    16

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command:HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-5. Site Retrieval Menu

1. ZIPCODE: The zip code for your study site. This will be translated into the latitude/longitude of the post office which services that zip code. This latitude/longitude will then be used as the center of the search area.

2. SITENAME: The name of the site for which the population is being retrieved.

3. STATE: The state identifier can be either the two-letter state postal abbreviation or the two-digit state Federal Information Processing Standard (FIPS) code.

4. DATA: The name of the Census database to be accessed for data. Your choices are:

   80POPULATION: for population figures from the 1980 Census
   80HOUSING: for housing figures from the 1980 Census
   83POPULATION: for population figures from the Donnelly Marketing estimates.

There are no data available on 1983 housing figures.

5. STANRING: A yes/no switch for standard ring distances. Leave this as YES to use the default ring distances. Enter NO to enter your own ring distances on a later menu. The default ring distances are: .5, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0, 15.0, 25.0 and 50.0 kilometers.

6. SECTORS: The number of sectors to divide the circle into. A default of 16 sectors is provided (16 sectors are the maximum allowed).

If you selected the DATASET input method from the first menu, then the menu shown below appears instead of the menu shown previously.

```
MENU: Census Data Retrieval

ref parmname    parameter description                  value  index
1.  IDATASET    Name of the input dataset
2.  ZIPVAR      Variable name which contains ZIPCODE
3.  NAMEVAR     Variable name - contains SITENAME
4.  STABVAR     Variable name - STATE abbreviation
5.  DATA        Name of the Census database
6.  STANRING    Use standard ring distances            YES
7.  SECTORS     Number of sectors                      16

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

FIGURE 4-6. Census Data Retrieval Menu

1. IDATASET: The name of the input dataset in the GEMS dataset library. These data must have the required geocode variables (in this case zip code).

2. ZIPVAR: The name of the variable in the input dataset which contains the zip codes with which retrievals will be done.

3. NAMEVAR: The name of the variable in the input dataset which contains a site name or text string to be used to identify the sites.

4. STABVAR: The name of the variable in the input dataset which contains the two-letter state postal abbreviation or a two-digit state Federal Information Processing System (FIPS) code.

5. DATA: The name of the Census database to be accessed for data. Your choices are:

80POPULATION:population figures from the 1980 Census
80HOUSING: housing figures from the 1980 Census
83POPULATION: population figures from the Donnelly Marketing estimates.

There are no data available on 1983 housing figures.

6. STANRING: A yes/no switch for standard ring distances. Leave this as YES to use the default ring distances. Enter NO to enter your own ring distances on a later menu. The default ring distances are: .5, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0, 15.0, 25.0 and 50.0 kilometers.

7. SECTORS: Enter the number of sectors to divide the circle into. A default of 16 sectors is provided (16 sectors are the maximum allowed). .

When you have filled in the required parameters on the appropriate menu above, enter [redacted] to proceed.

If you entered YES for the STANRING parameter, then continue below with the run mode specification. If you entered NO for the STANRING parameter, then you must enter the ring distances to be used for your retrieval. Enter the distances in kilometers when the following prompts appear:

Enter the first ring distance
GEMS >

This is followed by:

Enter the next ring distance
GEMS >

The second prompt will repeat. When you are through entering ring distances, you may enter a carriage return in response. The next step will depend on the options you selected on the first menu. If you selected the DIRECT input, the output will be displayed at your terminal. If you chose the dataset input option, then you will continue with:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

Enter your preferred method to run the program. If you select the interactive mode, the table will be displayed on your terminal. If you entered batch, GEMS will send the job to the batch queue and inform you of that fact. If running in batch, you must have selected a report or dataset output option or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. Afterwards you will be returned to the GEOLIST menu.

4.1.2 Access Census Data (CENSUS)

The CENSUS procedure retrieves the population within an area of interest. This area is defined through the specification of a center point and radius to define a circle. The circle is then broken down into sections by specifying sectors (segments defined by radial lines) and rings (sections defined by additional rings drawn within the major ring of the circle's

radius). For each sub-section defined by the above schema, the population will be reported. The totals within the sectors and rings and the overall total will also be displayed.

This procedure can accept input from either your terminal or from a GEMS dataset. The procedure can also be used to create an output GEMS dataset or report in your user dataset library. When you select this procedure, you will first set up the parameters in the parameter entry menu shown below:

MENU: INPUT/OUTPUT SPECIFICATION

ref	parmname	parameter description	value index
1.	METHOD	Input method specification	DIRECT
2.	DATASET	Name of the output dataset	BROWSE
3.	DTSTAG	Tag field of the output dataset	*
4.	REPORT	Name of the output report	BROWSE
5.	RPRTAG	Tag field of the output report	*

Enter one or more combinations of: reference or parameter name and value(s)[ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-7. Input/Output Specifications (CENSUS)

1. METHOD: The input method defines the source of the input to the procedure. The default is DIRECT which runs the procedure interactively with the input coming from your terminal. The other option is DATASET which indicates that the input is to come from a GEMS dataset. Note that if you select dataset input, you cannot select BROWSE for output. You must specify a dataset or report name, which directs the output to a GEMS dataset or report.

2. DATASET: The name of the output dataset. If the default of BROWSE is selected, then results appears on your terminal and no output dataset is created. If you want an

output dataset, then you must enter the name of the dataset you would like the output stored under in your user dataset library.

3. DSTAG: The tag field of the output dataset. The tag field is a 60 character field which can be used to determine the contents of a stored dataset.

4. REPORT: The name of the output report. If the default of BROWSE is selected, then results appear on your terminal and no output report is created. If you want an output report, then you must enter the name of the report to the name you would like the output stored under in your user dataset library.

5. RPTAG: The tag field of the output report. The tag field is a 60 character field which can be used to determine the contents of a stored report.

If you only wish to interactively view the data at the menu above, you do not have to make any changes. Once you have made the selections, enter [redacted] to move onto the next menu where you will specify the coordinate system to use to specify the center of the area of interest:

```
MENU: Select Method to Locate the Area of Interest

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude (decimal degrees)
4. Latitude/Longitude (degrees, minutes, seconds)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-8. Geographic Coordinate Selection Menu

Select the method you will be using to specify the coordinates for your area of interest. Only the menu that appears for the zip code will be shown since the options all display similar menus. For this example, the input method is DIRECT.

```
MENU: CENSUS Data Retrieval

ref parmname    parameter description                value  index
1. ZIPCODE      Zip code of the study site
2. SITENAME     Name of the study site
3. DATA         Name of the Census database
4. STANRING     Use standard ring distances          YES
5. SECTORS      Number of sectors                    16

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS>
```

FIGURE 4-9. Census Data Retrieval Menu (Direct Option)

1. ZIPCODE: The zip code for your study site. This will be translated into the latitude/longitude of the post office which services that zip code. This latitude/longitude will then be used as the center of the search area.

2. SITENAME: The name of the site for which the population is being retrieved.

3. DATA: The name of the Census database to be accessed for data. Your choices are:

   80POPULATION: population figures from the 1980 Census
   80HOUSING: housing figures from the 1980 Census
   83POPULATION: for population figures from the Donnelly Marketing estimates

   There are no data available on 1983 housing figures.

4. STANRING: A yes/no switch for standard rin distances. Leave this as YES to use th default ring distances. Enter NO to ente your own ring distances on a later menu The default ring distances are: .5, 1.0, 2.0 3.0, 4.0, 5.0, 10.0, 15.0, 25.0 and 50. kilometers.

5. SECTORS: The number of sectors to divide the circl into. A default of 16 is provided. maximum of 32 sectors is allowed.

If you selected the DATASET input method from the first menu the parameter entry menu shown below appears instead of th one shown above.

MENU: CENSUS Data Retrieval

ref	parmname	parameter description	value	index
1.	IDATASET	Name of the input dataset		
2.	ZIPVAR	Variable name which contains ZIPCODE		
3.	NAMEVAR	Variable name - contains SITENAME		
3.	DATA	Name of the Census database		
4.	STANRING	Use standard ring distances	YES	
5.	SECTORS	Number of sectors	16	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-10. Census Data Retrieval (Dataset Option)

1. IDATASET: The name of the input dataset in th GEMS dataset library. These data mus have the required geocode variables (i this case, zip code).

2. ZIPVAR: The name of the variable in the inpu dataset which contains the zip codes wit which retrievals will be done.

3. NAMEVAR: The name of the variable in the input dataset which contains a sitename or text string which can be used to identify the sites.

The remaining parameters are the same as described for the previous parameter entry menu.

When you have filled in the required parameters on the appropriate menu above, enter [redacted] to proceed.

If you entered YES for the STANRING parameter, then continue below with the run mode specification. If you entered NO for the STANRING parameter, you must enter the ring distances to be used for your retrieval. Enter the distances in kilometers when the following prompts appear:

Enter the first ring distance
GEMS >

This is followed by:

Enter the next ring distance
GEMS >

The second prompt will continue. When you are through entering ring distances, you may enter a [redacted] to end distance specification. You will then continue with:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

Enter your preferred method to run the program. If you select the interactive mode, then the table will be displayed on your terminal. If you entered batch, GEMS will send the job to the batch queue and inform you of that fact. If running in batch, you must have selected a report or dataset output option or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. Afterwards you will be returned to the GEOLIST menu.

4.1.3 Determine County Coverage (COVERAGE)

The COVERAGE procedure will, when given a point and the radius of a circle around that point, report all of the counties which have all or a portion of their land area within the region of interest. This procedure is very useful in determining the geographic coverage of an area which is to be studied.

This procedure can accept input from either your terminal or from a GEMS dataset. The procedure can also be used to create an output GEMS dataset or report in your user dataset library. When you select the COVERAGE procedure, the parameter entry menu shown below appears.

MENU: INPUT/OUTPUT SPECIFICATION

ref	parmname	parameter description	value index
1.	METHOD	Input method specification	DIRECT
2.	DATASET	Name of the output dataset	BROWSE
3.	DTSTAG	Tag field of the output dataset	*
4.	REPORT	Name of the output report	BROWSE
5.	RPRTAG	Tag field of the output report	*

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-11. Input/Output Specifications (COVERAGE)

1. METHOD: The input method which defines the source of the input to the procedure. The default is DIRECT which runs the procedure interactively with the input coming from your terminal. The other option is DATASET which indicates that the input is to come from a GEMS dataset. If you select dataset input, you cannot select BROWSE for output. You must specify a dataset or report name, which directs the output to a GEMS dataset or report.

2. DATASET: The name of the output dataset. If the default of BROWSE is selected, then results appear on your terminal and no output dataset is created. If you want an output dataset, you must enter the name of the dataset you would like the output stored under in your user dataset library.

3. DSTAG: The tag field of the output dataset. The tag field is a 60 character field which can be used to determine the contents of a stored dataset.

4. REPORT: The name of the output report. If the default of BROWSE is selected, no output report is created. If you want an output report, enter a name.

5. RPTAG: The tag field of the output report. The tag field is a 60 character field which can be used to determine the report contents.

If you only wish to interactively view the data, you do not have to make any changes at the above menu. Once you have made your selections, you should enter [illegible] to move onto the next menu where you will specify the coordinate system to use to specify the center of the area of interest.

```
MENU: Select Method to Locate the Area of Interest

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude (decimal degrees)
4. Latitude/Longitude (degrees, minutes, seconds)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-12. Geographic Coordinate Selection Method

Select the method you will be using to specify the coordinates for your area of interest. Only the menu that appears for the

zip code will be shown since the menus that appear for the various methods are all similar except for the geographic code specification. For this example, the input method is DIRECT.

```
MENU: Find Geographic Coverage by Zip code

ref parmname   parameter description          value   index
1. ZIPCODE ,   Zip code of site
3. RADIUS      Radius in Kilometers             0

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-13. Zip Code Entry Menu

1. ZIPCODE: The zip code of the site for which you are searching.

2. RADIUS: The radius of the circle to use as the area of interest. If you leave the default of zero (0), then only the county containing the specified point will be displayed.

If you had selected the dataset input option, the following menu will appear.

```
MENU: Find Geographic Coverage by Zip code

ref parmname   parameter deecription          value   index
1. IDATASET    Name of the input dataset
2. ZIPVAR      Variable which contains ZIPCODE
3. RADIUS      Radius in Kilometers             0

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

FIGURE 4-14. Find Geographic Coverage by Zip Code Menu

1. IDATASET: The name of the dataset to be used as input to the procedure.

2. ZIPVAR: The variable in the input dataset which has as its value the zip code of the site of interest.

3. RADIUS: The radius to be used to define the circle which is the area of interest. If the default of zero (0) is selected, then only the county which contains the point of interest will be located.

Had you selected a geographic location method other than zip code as the way to locate the search point, the geographic location variables above would have been replaced by variables appropriate to the search method selected.

Once you have filled in the appropriate menu, enter **NEXT** to continue to the run specification.

Enter program execution mode: B (batch) or I (interactive)
GEMS >

Enter your preferred method to run the program. If you select the BROWSE option and the interactive mode, the procedure will print out the counties it has found. If you entered batch, then GEMS will send the job to the batch queue and inform you of that fact. If running in batch, you must have selected a report or dataset output option or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. Afterwards you will be returned to the GEOLIST menu.

4.1.4 Geographic Data Management (GEODM)

The Geographic Data Management (GEODM) procedure allows you to extract data from a GEMS dataset for an area of interest. Data is retrieved by entering a location and radius, and then by processing the dataset for user selected variables. You may also create new variables during the extraction by

modifying old variables using various formula, combining tv or more variables, or in other ways manipulating existir variables. The dataset must have the latitudes and longitud of the sites stored as variables. The procedure allows you define an area of interest using a central point and a radi within which the variables in the GEMS dataset can b manipulated. You can either create a subset of the origin dataset which contains only the variables which you want f the area of interest, or you can have GEMS calculate ne variables based on the information contained in the datas being manipulated.

When you enter the GEODM procedure, the followin parameter entry menu appears.

MENU: Geographic Data Management

ref	parmname	parameter description	value	index
1.	INSET	Name of the input dataset		
2.	METHOD	Method of Study Site Identification		
3.	TABLE	Title for the output table		
4.	OUTSET	Name of the output dataset		
5.	TAG	Tag field for the output dataset	*	
6.	MAP	Automatic map creation flag	NO	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2, ..] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-15. Geographic Data Management Menu

1. INSET: The name of the dataset in the GEMS dataset library which is to be used as input to the procedure.

2. METHOD: The input location identification method. Enter either LAT/LONG (latitude/longitude) or ZIP (zip code).

3. TABLE: The title you would like to use for the table that the procedure creates.

4. OUTSET: The name to be used for the new output dataset which will be created in your GEMS dataset library.

5. TAG: The tag field for the output dataset. This is a descriptive field of up to 60 characters describing the contents of the dataset.

6. MAP: A yes/no switch giving you the option to automatically create a map using the data which you are manipulating with the GEODM procedure.

Once you have entered the information necessary in the above menu, enter NEXT to proceed to the next step. Remember that both the dataset name and table name must be entered since they both will be created. The prompts you will see will depend on the method of site location you specified above.

If you specified zip code as your location method, you will see the following prompt:

Enter the zip code of the study site
GEMS >

Here you enter the zip code you are using to locate your study site.

If you specified the lat/long location method, the menu shown below appears.

MENU: Tabulation Variables

ref	parmname	parameter description	value	index
1.	LATITUDE	Latitude of the study site		
2.	LONGITUDE	Longitude of the study site		

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT, BACK,END,CLEAR,EXIT
GEMS >

FIGURE 4-16. Tabulation Variables Menu

1. LATITUDE: The latitude of the study site. You may specify the latitude in either degrees, minutes and seconds (DDMMSS) or decimal degrees.

2. LONGITUDE: The longitude of the study site. You may specify the longitude as either DDMMSS or decimal degrees.

Please note that the UTM coordinate system is not available in this procedure. Once you have identified the site location, you must specify the radius to be used to identify the area of interest:

Enter the study radius in km
Default = 0
GEMS >

After entering the radius, you now define the variables to be used in the tabulation:

Enter the name of the first tabulation variable
GEMS >

You may enter the name of any existing variable in the GEMS dataset being used for input. This will add that variable to the table and the new dataset being created by this procedure.

You may also define the range of the variable or equation. This gives you the ability to extract variables only when the value falls within the specified range. An example of this is "variable [1:100]" which would have the effect of extracting the specified variable only if its value fell within 1 and 100. You can also create a new variable with this procedure using variables existing in the dataset already in an equation.

Example: newvar = oldvar1*(oldvar2 + oldvar10)

The mathematical operators allowed in an equation are: plus (+), minus (-), multiplication (*), exponentiation (**), common logarithm (log10(variable name)), natural logarithm (log) or

concatenation (//). You would enter this equation in response to the prompt given above.

When creating a new variable, the default for numeric variables is to create the new variable as a real variable if any variable in the equation is real; otherwise the new variable is created as an integer. You can force the new variable to be real by multiplying the equation by 1.0.

After specifying a variable, you will then be asked:

Enter YES or NO to include this variable in the table
Default = YES
GEMS >

If you enter [redacted], the variable will be displayed in the table which the procedure will create. The procedure will continue with the following:

Enter the name of the next tabulation variable
GEMS >

This prompt will repeat until you enter a [redacted] for the above prompt. That will signal an end to your input of data. When you signal an end to your input of data, the procedure will create the new dataset and display the table of values at your terminal. When the procedure has finished, it will automatically enter the Draw County Map procedure if you selected YES for the Automatic Map Creation flag. For documentation on this procedure, see section 4.2.2. If you selected NO for the map creation flag, then you will return to the GEODM main menu. Enter [redacted] if you wish to return to the GEOLIST menu.

4.1.5 HUCODE/SOIL locator (HUCODE)

The HUCODE procedure provides identification of the U.S. Geological Survey (USGS) hydrologic unit code for the region of the study site. Hydrologic unit boundaries are of environmental significance as they define land areas of similar hydrologic properties.

This procedure can accept input from either your terminal or from a GEMS dataset. The procedure can also be used to create an output GEMS dataset or report in your user dataset library. When you select this procedure, you will first set up the parameters in the following parameter entry menu.

MENU: INPUT/OUTPUT SPECIFICATION

ref	parmname	parameter description	value	index
1.	METHOD	Input method specification	DIRECT	
2.	DATASET	Name of the output dataset	BROWSE	
3.	DTSTAG	Tag field of the output dataset	*	
4.	REPORT	Name of the output report	BROWSE	
5.	RPRTAG	Tag field of the output report	*	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-17. Input/Output Specifications Menu (HUCODE)

1. METHOD: The input method which defines the source of the input to the procedure. The default is DIRECT which runs the procedure interactively with the input coming from your terminal. The other option is DATASET which indicates that the input is to come from a GEMS dataset. If you select dataset input, you cannot select BROWSE for output. You must specify a dataset or report name, which directs the output to a GEMS dataset or report.

2. DATASET: The name of the output dataset. If the default of BROWSE is selected, then results appear on your terminal and no output dataset is created. If you want an output dataset, you must enter the name of the dataset you would like the output stored under in your user dataset library.

3. DSTAG: The tag field of the output dataset. The tag field is a 60 character field which can be used to determine the contents of a stored dataset.

4. REPORT: The name of the output report. If the default of BROWSE is selected, then the results appear on your terminal and no output report is created. If you want an output report, then you must enter the name of the report you would like the output stored under in your user dataset library.

5. RPTAG: The tag field of the output report. The tag field is a 60 character field which can be used to determine the contents of a stored report.

If you only wish to interactively view the data at the menu above, do not make any changes. Enter NEXT to move onto the next menu, where you will specify the coordinate system to use to specify the center of the area of interest.

```
MENU: Select Method to Locate the Area of Interest

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude (decimal degrees)
4. Latitude/Longitude (degrees, minutes, seconds)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-18. Geographic Coordinate Selection Menu

Select the method you will be using to specify the coordinates for your area of interest. The menus that appear subsequently are similar except for the geographic location parameters. Therefore, only one example will be shown. For the purposes of this example, we are assuming that zip code was selected. Also, the input method is DIRECT.

```
MENU: SOIL/HUCODE data retrieval by Zip code

ref  parmname   parameter description                  value   index
1.  STATE       State identifier
2.  ZIPCODE     Zip code of a data location

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-19. SOIL/HUCODE Data Retrieval Menu

1. STATE: The state identifier. It can be either the two-letter state postal abbreviation or the two-digit state FIPS code.

2. ZIPCODE: The zip code of the study site.

If you had chosen DATASET as the input mode, the following menu would have appeared.

```
MENU: SOIL/HUCODE data retrieval by Zip code

ref  parmname   parameter description                  value   index
1.  IDATASET    Name of the Input Dataset
2.  STABVAR     Variable name -STATE abbrev.
3.  ZIPVAR      Variable name- ZIPCODE

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-20. SOIL/HUCODE Data Retrieval (Dataset Option)

1. IDATASET: The name of the dataset to be used as input to the HUCODE procedure.

2. STABVAR: The variable in IDATASET which identifies the state. It may be either the two-digit

state FIPS code or the two-letter postal abbreviation for the state.

3. ZIPVAR: The variable in IDATASET which has the site zip code as its value.

After entering the information, enter NEXT to proceed to the next step where you will be asked for how you would like to execute the procedure:

Enter program execution mode: B (batch) or I (interactive)
GEMS>

If you enter B, the program will be submitted to the batch queue. Remember that in order to run in batch, you have to be directing the output to a GEMS dataset or report or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. If you enter I, the procedure will run and display at your terminal.

4.1.6 Convert Geocode to Lat/Long (LATLON)

This procedure allows you to convert one coordinate system to latitude/longitude coordinates. You can enter a zip code or UTM coordinate to retrieve the latitude/longitude which GEMS would use for it. When you enter a zip code, the Zipcode database, which consists of the post offices in the U.S., is searched. If found, the latitude/longitude will be displayed. Note that not all zipcodes are in this database, as new ones have been created since this database was implemented. Also, the name of the town, if available, will also be displayed. For UTM coordinates, a conversion program acquired from the U.S. Geological Service is used to translate the UTM coordinates to latitude/longitude.

You may also use this procedure to determine where in the United States, a given latitude/longitude is. If you enter a latitude/longitude, the procedure will search the Zipcode database to find the nearest post office to your latitude/longitude. When found, the procedure will display the name, state, county, FIPS code and latitude/longitude of the

post office. The distance from your entered latitude/longitude to the located post office will also be displayed.

This procedure can accept input from either your terminal or from a GEMS dataset. The procedure can also be used to create an output GEMS dataset or report in your user dataset library. When you select the LATLON procedure, you will first set up the parameters in the following parameter entry menu.

```
MENU: INPUT/OUTPUT SPECIFICATION

ref parmname   parameter description          value  index
1.  METHOD     Input method specification     DIRECT
2.  DATASET    Name of the output dataset     BROWSE
3.  DTSTAG     Tag field of the output dataset  *
4.  REPORT     Name of the output report      BROWSE
5.  RPRTAG     Tag field of the output report  *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS>
```

FIGURE 4-21. Input/Output Specification Menu (LATLON)

1. METHOD: The input method which defines the source of the input to the procedure. The default is DIRECT which runs the procedure interactively with the input coming from your terminal. The other option is DATASET which indicates that the input is to come from a GEMS dataset. Note that if you select dataset input, you cannot select BROWSE for output. You must specify a dataset or report name, which directs the output to a GEMS dataset or report if you select the DIRECT input method.

2. DATASET: The name of the output dataset. If the default of BROWSE is selected, no output dataset is created. If you want an output dataset, you must enter the name of the

dataset you would like the output stored under in your user dataset library.

3. DSTAG: The tag field of the output dataset. The tag field is a 60-character field which can be used to determine the contents of a stored dataset.

4. REPORT: The name of the output report. If the default of BROWSE is selected, then the results appear on your terminal and no output report is created. If you want an output report, enter the name of the report you would like the output stored under in your user dataset library.

5. RPTAG: The tag field of the output report. The tag field is a 60-character field which can be used to determine the contents of a stored report.

The above menu has been defaulted so that if you only wish to interactively view the data you do not have to make any changes. To move onto the next menu when you have made your selections, enter [illegible]. The following menu appears.

```
MENU: Select Method to Locate the Area of Interest

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude (decimal degrees)
4. Latitude/Longitude (degrees, minutes, seconds)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-22. Geographic Coordinate Selection Menu

Select the method you will be using to specify the coordinates for your area of interest. The menus that appear subsequently are similar except for the geographic location parameters.

Therefore, only one example for the zip code option will be shown.

```
MENU: Zip code locator

ref parmname    parameter description                    value   index
1.  ZIPCODE     Zip code of the location

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-23. Zip Code Locator Menu

1. ZIPCODE: The zip code for which to retrieve the location information.

If you had entered DATASET as the input method, the menu you would see instead is shown below.

```
MENU: Zip code locator

ref  parmname   parameter description                    value   index
1.  IDATASET    Name of the input dataset
1.  ZIPVAR      Variable name which contains ZIPCODE

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-24. Zip Code Locator Menu (Dataset Option)

1. IDATASET: The name of the dataset in the GEMS dataset library which contains the variable to be used as input to the procedure.

2. ZIPVAR: The name of the variable in the input dataset which contains the zip codes to be processed.

After entering the information, enter NEXT to proceed to the next step where you will be asked:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

If you enter B, the program will be submitted to the batch queue. Remember, in order to run in batch, you have to be directing the output to a GEMS dataset or report output option or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. If you enter I, then the procedure will run and display the retrieved information at your terminal.

4.1.7 Lookup/Examine STAR Station Data (STAR)

The STAR procedure locates the closest STAR stations to a specified point. The procedure will always display at least three STAR stations (up to a maximum of seven STAR stations) nearest the specified point. (These data are automatically accessed and used by some of the air models in GEMS to determine atmospheric conditions to use when modeling.) For a description of the data itself, refer to Chapter 10 on the STAR dataset.

This procedure can accept input from either your terminal or from a GEMS dataset. The procedure can also be used to create an output GEMS dataset or report in your user dataset library. When you select the STAR procedure, you will first set up these parameters in the parameter entry menu below.

```
MENU: INPUT/OUTPUT SPECIFICATION

ref parmname   parameter description          value    index
1. METHOD      Input method specification     DIRECT
2. DATASET     Name of the output dataset     BROWSE
3. DTSTAG      Tag field of the output dataset *
4. REPORT      Name of the output report      BROWSE
5. RPRTAG      Tag field of the output report *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END,CLEAR,EXIT
GEMS >
```

FIGURE 4-25. Input/Output Specifications Menu (STAR)

1. METHOD: The input method defines the source of the input to the procedure. The default is DIRECT which runs the procedure interactively with the input coming from your terminal. The other option is DATASET which indicates that the input is to come from a GEMS dataset. You must specify a dataset or report name, which directs the output to a GEMS dataset or report.

2. DATASET: The name of the output dataset. If the default of BROWSE is selected, then results appear on your terminal and no output dataset is created. If you want an output dataset, enter the name of the dataset you would like the output stored under in your user dataset library.

3. DSTAG: The tag field of the output dataset. The tag field is a 60 character field which can be used to determine the contents of a stored dataset.

4. REPORT: The name of the output report. If the default of BROWSE is selected, then the results appear on your terminal and no

output report is created. If you want an output report, enter the name of the report you would like the output stored under in your user dataset library.

5. RPTAG: The tag field of the output report. The tag field is a 60 character field which can be used to determine the contents of a stored report.

If you only wish to interactively view the data, you do not have to make any changes at the menu above. To move onto the next menu, enter NEXT. The following parameter entry menu appears.

```
MENU: Select Method to Locate the Area of Interest

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude (decimal degrees)
4. Latitude/Longitude (degrees, minutes, seconds)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-26. Geographic Coordinate Selection Menu

Select the method you will be using to specify the coordinates for your area of interest. The menus that appear subsequently are similar except for the geographic location parameters. Therefore, only the zip code method will be shown. Also, the input method is DIRECT.

```
MENU: Zip code locator

ref parmname   parameter description          value   index
1. ZIPCODE     Zip code of the location

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

FIGURE 4-27. Zipcode Locator Menu (STAR)

1. ZIPCODE: The zip code of your site.

If DATASET is the input method, the following menu appears.

```
MENU: Zip code locator

ref parmname   parameter description                value   index
1. IDATASET    Name of the input dataset
1. ZIPVAR      Variable name which contains ZIPCODE

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

FIGURE 4-28. Zip Code Locator Menu (Dataset Option)

1. IDATASET: The name of the dataset in the GEMS dataset library which contains the variable to be used as input to the procedure.

2. ZIPVAR: The name of the variable in the input dataset which contains the zip codes to be processed.

After entering the information, enter NEXT to proceed to the next step where you will be asked:

Enter program execution mode: B (batch) or I (interactive)
GEMS

If you enter B, the program will be submitted to the batch queue. Remember that in order to run in batch, you must be directing the output to a GEMS dataset or report or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. However, if you enter I, then the procedure will run and display at your terminal.

4.1.8 Find US Cities (USCITY)

The USCITY procedure allows you to retrieve information on the location of many cities in the United States. You may use

this procedure to locate a specific city or to retrieve all cities in the database within a specified area. This procedure can accept input from either your terminal or from a GEMS dataset. The procedure can also be used to create an output GEMS report in your user dataset library. When you select the USCITY procedure, the following menu appears.

```
MENU: INPUT/OUTPUT SPECIFICATION

ref parmname   parameter description          value    index
1. METHOD      Input method specification     DIRECT
2. DATASET     Name of the output dataset     BROWSE
3. DTSTAG      Tag field of the output dataset *
4. REPORT      Name of the output report      BROWSE
5. RPRTAG      Tag field of the output report *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

FIGURE 4-29. Input/Output Specification Menu (USCITY)

1. METHOD: The input method defines the source of the input to the procedure. The default is DIRECT which runs the procedure interactively with the input coming from your terminal. The other option is DATASET which indicates that the input is to come from a GEMS dataset. If you select dataset input, you cannot select BROWSE for output. You must specify a dataset or report name, which directs the output to a GEMS dataset or report.

2. DATASET: The name of the output dataset. If the default of BROWSE is selected, then results appear on your terminal screen and no output dataset is created. If you want an output dataset, you must enter the name of the dataset you would like the output stored under in your user dataset library.

3. DSTAG: The tag field of the output dataset. The tag field is a 60 character field which can be used to determine the contents of a stored dataset.

4. REPORT: The name of the output report. If the default of BROWSE is selected, then results appear on your terminal and no output report is created. If you want an output report, enter the name of the report you would like the output stored under in your user dataset library.

5. RPTAG: The tag field of the output report. The tag field is a 60 character field which can be used to determine the contents of a stored report.

The above menu has been defaulted so that if you only wish to interactively view the data, you do not have to make any changes. To move to the next menu where you will choose how to limit the search, enter NEXT

```
MENU: Select method to search for US cities

1. All cities in a state
2. Selected county in a specified state
3. Selected city in a specified state
4. ALL cities within a specified radius

Enter an option number or a command: HELP,BACK,END,CLEAR,EXIT
GEMS>
```

FIGURE 4-30. Geographic Coordinate Selection Menu

Each of these methods will be described below.

4.1.8.1 All Cities in a State

This option will display all cities in the database which are in a specified state. If you select this option, the following menu will appear.

```
MENU: Retrieve a list of all cities in a state

ref  parmname   parameter description                value   index
1.   STATE      State abbreviation or the state FIPS


Enter one or more combinations of: reference or parameter name and
value(s)[ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-31. City List Retrieval Menu

1. STATE: The two-letter postal abbreviation for the state or the 2-digit state FIPS code for the state.

After entering the information, enter **NEXT** to proceed to the next step where you will be asked:

Enter program execution mode: B (batch) or I (Interactive)
GEMS >

If you enter **B**, the program will be submitted to the batch queue. Remember that in order to run in batch, you must be directing the output to a GEMS dataset or report or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. If you enter **I**, then the procedure will run and display at your terminal.

4.1.8.2 Selected County in a Specified State

This procedure is the same as that in section 4.1.8.1 except that you must enter a three-digit FIPS county code also. The parameter entry menu below appears if you select this option.

```
MENU: Retrieve cities in a specified county of a specified state

ref parmname   parameter description          value   index
1. STFIPS      State FIPS code
2. COFIPS      County FIPS code

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

FIGURE 4-32. Retrieval of Cities within a County Menu

1. STFIPS: The two-digit state FIPS code for the state you would like to search.

2. COFIPS: The three-digit FIPS county code for the county to display to cities for.

After entering the information, enter NEXT to proceed to the next step where you will be asked:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

If you enter B, the program will be submitted to the batch queue. Remember to run in batch you have to be directing the output to a GEMS dataset or report or both. Also, if you are using the dataset input option, no output will be displayed on the terminal.If you enter I, then the procedure will run and display at your terminal.

4.1.8.3 Selected City in a Specified State

If you select this option, the following parameter entry menu appears. It is followed by a discussion of each parameter.

```
MENU: Search for a specified city in a specified state

ref parmname   parameter description              value   index
1. STATE       State abbreviation or state FIPS
2. CITY        City name

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS>
```

FIGURE 4-33. City Search Menu

1. STATE: The two-letter postal abbreviation for the state or the two-digit state FIPS code for the state.

2. CITY: The name of the city you would like to find. You may enter an asterisk (*) at the end of a portion of a city name to search for all cities beginning with a specified string. Example: NEW* would retrieve all cities whose names begin with NEW in that state.

After entering the information, enter NEXT to proceed to the next step where you will be asked:

Enter program execution mode: B (batch) or I (interactive)
GEMS>

If you enter B, the program will be submitted to the batch queue. Remember that in order to run in batch, you have to be directing the output to a GEMS dataset or report or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. If you enter I, then the procedure will run and display at your terminal.

4.1.8.4 All Cities Within a Specified Radius

This procedure will retrieve all cities which are within a certain area of interest. This area is defined by a central point and a radius to define a circle. The first step is to define the coordinate system to be used to specify the central point.

```
MENU: Select Method to Locate the Area of Interest

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude (decimal degrees)
4. Latitude/Longitude (degrees, minutes, seconds)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-35. Geographic Coordinate Selection Menu

Select the method you will be using to specify the coordinates for your area of interest. The menus that appear subsequently are similar except for the geographic location parameters. Therefore, only one example (zipcode) will be shown. Also, the input method is DIRECT.

```
MENU: Find cities using a central point and radius

ref parmname   parameter description          value   index
1. ZIPCODE     Zip code of a city
2. RADIUS      Search radius

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-34. City Search Using Radius Menu

1. ZIPCODE: The zip code which will be used as the central point for the search.

2. RADIUS: The radius in kilometers of the area to be searched.

After entering the information, enter NEXT to proceed to the next step where you will be asked:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

Remember that in order to run in batch mode, you must be directing the output to a GEMS dataset or report or both. Also, if you are using the dataset input option, no output will be displayed on the terminal. If you select the interactive mode, then the procedure will run and display at your terminal.

4.2 Geodata Handling Mapping Procedures

The GEMS mapping procedures give you a number of ways to display data contained in GEMS datasets. These maps allow you a great deal of choice as to type, size, and other display options. When creating a map, do not attempt to display too much data in one map for that will make the map extremely difficult to read. Note that it may take several tries to get a map which displays the data you want in a way you are happy with.

In order to access this operation, you either select option 2 from the Geodata Handling Menu, or enter the direct command GEOMAP. The following menu appears.

```
MENU: Geodata Handling Mapping Procedures

1. Draw Block Group/Enumeration Dist. Map      (BGEDMAP)
2. Draw County Map                             (CNTYMAP)
3. Hydrological Unit Map                       (HUMAP)
4. State/County Map                            (STCOMAP)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-36. GEOMAP Main Menu

4.2.1 Draw Block Group/Enumeration Dist. Map (BGEDMAP)

The BGEDMAP procedure draws the distribution of Census Block Group/Enumeration Districts (BG/EDs) around a user-identified point. You may specify the point using a number of geographic coordinates. You then specify the radius of a circle to be drawn around the selected point. The centroids will then be plotted for each BG/ED which is located within the defined circle.

You will first select the method to be used to specify the geographic location you want to map.

```
MENU: Select Method to Locate the Area of Interest

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude (decimal degrees)
4. Latitude/Longitude (degrees, minutes, seconds)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-37. Geographic Coordinate Selection Menu

The menu following the above menu will be specific to the method of geographic location specification you have chosen. If you choose the zip code option, then the menu will ask for the zip code. Since they are all similar, except for the geographical coordinates that are asked for, only the menu for option 4 will be illustrated.

```
MENU: Generate BGEDMAP by Lat/Long (degrees, minutes, seconds)

ref parmname    parameter description                  value   index
1.  LAT         latitude of the location(DDMMSS)
2.  LONG        longitude of the location(DDMMSS)
3.  MAP         name of the output map
4.  TITLE       title of the output map
5.  TAG         tag field of the output map            *
6.  RINGDIST    ring distances in Km
7.  NSECTORS    number of sectors
8.  CHAR        plot symbol on map                     x

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

FIGURE 4-38. Generate BG/ED Map Menu

1. LATITUDE: The latitude of the site around which to retrieve BG/ED centroids, specified in DDMMSS format.

2. LONGITUDE: The longitude of the site around which to retrieve BG/ED centroids, specified in DDDMMSS format

3. MAP: The name of the map to be created in your user dataset library.

4. TITLE: The title to be used to identify the map. This title will be displayed at the top of the map.

5. TAG: A field of up to 60 characters which can be used to describe the contents of the map for future reference.

6. RINGDIST: An array of numbers which defines the rings for which BG/ED centroids are to be displayed. This is an array variable which can be used to segment the table which is produced with the map. The table will

print out population numbers based on the ring which the BG/ED centroid falls into.

7. NSECTORS: The number of sectors to divide the map into. This can help in locating an area on the map.

8. CHAR: The plot character to be used to display the BG/ED centroids.

Once the above parameter entry menu has been filled in, you may enter NEXT to move onto the next step where you will be asked whether to run the procedure in batch or interactively:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

If you choose to run the procedure in batch mode, the job will run in the background and your terminal will be freed up for other uses. You will get a message that the job has been submitted and then returned to the GEOMAP menu. If you run the program interactively, the map will be displayed at your terminal. When the map is completed, the terminal bell will be rung to indicate completion. Enter a carriage return in order to clear the screen and return to the GEOMAP menu. Maps created either interactively or in batch will be stored in GEMS, for later retrieval. An example of a BG/ED map is shown in Figure 4-39.

4.2.2 Draw County Map (CNTYMAP)

The Draw County Map procedure allows you to produce maps based on the geographic distribution of data around a site.

At the main menu of this procedure, you set up the parameters which define the capabilities you will use in drawing the map. Input for the county map procedure falls into three general categories: location specification, overlay description including the drawing of stream reaches, and shading control. The location specification consists of either specifying the state

The plot center is at:

Latitude	38.8700
Longitude	77.0700

Cumulative Population	Within Radius
3334668	50.0

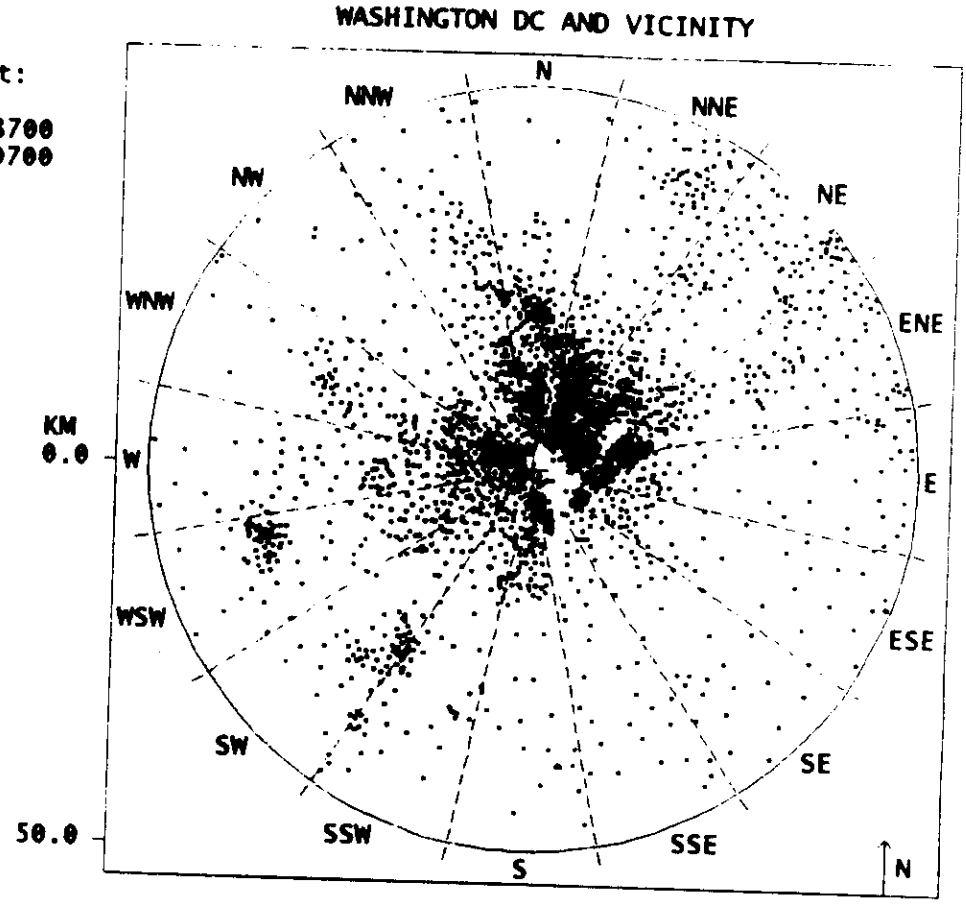

FIGURE 4-39. Example of BG/ED Map

or states to be mapped or entering a center and radius to define the area which is to be mapped. An overlay is an additional point or line data which is drawn on the map to display more information. The overlays possible with this procedure are: manual overlay by which method you enter a location and name to be drawn on the map, weather station overlays which display weather stations from the STAR station dataset, GEMS overlays which allow overlays to be done from any GEMS' dataset which contains the proper geocode variables, and river reach overlays which displays data from the Reachtrace database. A final option is county shading which allows you to fill the area within a county boundary with a shading pattern.

MENU: Draw County Map

ref	parmname	parameter description	value	index
1.	MAPNAM	Name of the output map dataset		
2.	HEADING	Heading of the map		
3.	TAG	Tag field of the output map		
4.	RING	Draw map rings		
5.	COUNTY	Display only specific counties		
6.	MANUALLY	Overlay a place manually		
7.	WEATHER	Weather station overlay		
8.	GEMSDATA	Gems dataset overlay		
9.	REACH	River reach overlay		
10.	SHADE	Shade specific counties		

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-40. Draw County Map Menu

1. MAPNAM: The name of the map to be created. The map will be stored under this name in your GEMS user library. This name can be up to 20 characters in length.

2. HEADING: The map title, to be displayed at the top of the map. This is a character string of up to 30 characters. Use this parameter to identify the area of the map.

3. TAG: A descriptive field of up to 60 characters. Enter a string which describes the content of the map to help you recall its contents.

4. RING: A yes/no switch which defines whether to draw rings on the map at specified distances. Enter Y to draw these rings or leave it as N not to. If you want to draw the rings you will be asked later for the distances at which to draw the rings.

5. COUNTY: A yes/no switch which allows you the option of drawing only specific counties in the area defined. If you enter a Y for this parameter, you will be prompted later for the FIPS codes of the counties to be drawn. Enter N (the default) to have all counties in the defined area be drawn.

6. MANUALLY: A yes/no switch to manually overlay sites. Enter Y to manually overlay sites on the map. Accept the default, N if you have no sites to overlay. This option allows you to identify points to be drawn with descriptive text for those points.

7. WEATHER: A yes/no switch to have the weather stations from the STAR station database drawn on the map. Enter YES for this parameter if you would like the STAR stations displayed. GEMS will automatically retrieve them from the STAR dataset.

8. GEMSDATA: A yes/no switch to overlay sites from a GEMS dataset. The dataset must have the latitude and longitude of the point and a character variable with descriptive text about the site. You will be prompted for this information in a later menu.

9. REACH: A yes/no switch to overlay river reaches. The stream reach data will show river locations on your map. There are a number of options available with this function which you can use to display all or some subset of the available data.

10. SHADE: A yes/no switch to shade counties based on selected data values.

Once you have selected the options you would like to use for your map, enter NEXT to continue to the next step in the process. The following menus may not all be used for every map; each option above has its own menu which is used to specify its parameters. The following combination of menus and prompts appear if you had entered Y to all the options in the menu above.

If you selected the option to overlay river reaches, the following menu is displayed:

MENU: Reach Display Criteria

ref	parmname	parameter description	value	index
1.	PLOTLVLS	Reach levels	*	(1)
2.	SHORES	Plot shore lines	YES	
3.	CATUNTS	Reach Cat-Unit no.	*	(1)
4.	LOGIC	Logical option (AND,OR)	AND	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS>

FIGURE 4-41. Reach Display Criteria Menu

1. PLOTLVLS: A subset of reaches may be drawn by specifying the "levels" which are to be drawn. The level of a reach are defined by the following rules:

- A reach level is 1 if and only if it flows into a major body of water.

- A reach level is some other number N which is greater than one (1) if and only if it flows into another reach of level N-1 (e.g. a level 2 reach can only have level 3 reaches flowing into it, and a level 3 reach can only flow into a level 2 reach).

2. SHORES: A yes/no switch which allows you to display non-transport reaches such as lakes, inlets, bays, etc. on the map. This option is defaulted to YES; enter a NO if you do not want this data displayed.

3. CATUNTS: The cataloging-unit numbers of the reaches you want to draw. Up to fifty cataloging-unit numbers can be entered. Enter the first six digits of a number to specify an entire river basin, or the first eight digits to indicate the inclusion of a sub-basin. To request the display of a particular reach, enter all 11 digits.

4. LOGIC: Defines the logic to be used in reach selection if you have specified both options 1 and 3 in this menu. AND specifies that the reach in question must meet both criteria (PLOTLVLS and CATUNTS); whereas OR specifies that all reaches which meet one or the other of the criteria will be plotted.

Once you have filled in this menu, enter NEXT to continue. (Note that the default for this menu is to draw all reaches in the display area.) The following prompt appears.

Enter map identification option (STATE or SITE)
GEMS >

Enter STATE if you would like to map entire states. Enter SITE in order to specify a central location and a radius within which to map.

If you enter STATE, then GEMS responds with:

Enter a state abbreviation
GEMS >

Enter the two-letter state abbreviation for the state you wish to work with. GEMS will then prompt you for another state abbreviation. Continue entering state abbreviations until you have entered all the states in which you are interested. In order to proceed to the next step, enter [redacted]

If you had entered SITE you would be asked:

Enter study site location method (ZIPCODE or LAT/LONG)
GEMS >

If you answered ZIPCODE, the following prompt is displayed:

Enter the zip code of the study site
Default = 0
GEMS >

Enter the zip code whose latitude/longitude coordinate is to be used as the central point for the map.

If you answer LAT/LONG, the following menu will appear.

```
MENU: Define County Map Location

ref parmname    parameter description                value    index
1. LATITUDE     Latitude of the study site
2. LNGITUDE     Longitude of the study site

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command: HELP, NEXT,
BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-42. Define County Map Location Menu

1. LATITUDE: The latitude of the study site in either DDMMSS format or as decimal degrees

2. LNGITUDE: The longitude of the study site in either DDDMMSS format or as decimal degrees.

Once you have specified the coordinates for the center of your map, enter NEXT to continue.

Once you have specified the center point of the map, you will then be asked to enter the radius to be used to define the area.

Enter the study radius (km)
Default = 0
GEMS >

Once you have entered the study radius, you will be asked to specify the distance to be used for the rings to be drawn on the map (selected by an option from the first menu).

Enter radius (km) for the first ring
Default = 0
GEMS >

GEMS continues with:

Enter the radius (km) for the next circle
Default = 0
GEMS >

Once you have entered all of the distances at which you would like to have rings drawn, you may respond to this prompt with just a Carriage return . GEMS will then display the prompts which allow you to select a subset of counties within the selected area for display, as shown in the example below:

The selected states are:
NY FIPS = 36
Enter a 5-digit state and county FIPS
GEMS >

Here enter a state and county FIPS code for the county which you would like to have plotted. This prompt will repeat to allow you to specify additional counties to be plotted until you respond with just a carriage return to signify that all counties have been entered.

The next menu to be displayed is the menu to manually overlay a place.

```
MENU: OVERLAY A PLACE MANUALLY

ref parmname    parameter description          value   index
1. METHOD       Location method
2. ZIPCODE      Zip code of overlay site
3. LATITUDE     Latitude of overlay site
4. LNGITUDE     Longitude of overlay site
5. PLACE        Overlay place name             *
6. ANOTHER      Overlay another place?         YES

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2, ...] or a command:HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS>
```

FIGURE 4-43. Manual Overlay Menu

1. METHOD: The site location method. Enter ZIPCODE or LAT/LON. If you enter ZIPCODE then you must enter a value for item two (2) below. If you enter LAT/LON, you must enter values for items three (3) and four (4).

2. ZIPCODE: The zip code which is to be used to locate the overlay site.

3. LATITUDE: The latitude of the overlay site. You may enter the latitude in either DDMMSS format or as decimal degrees.

4. LNGITUDE: The longitude of the overlay site. You may enter the longitude in either DDMMSS format or as decimal degrees.

5. PLACE: The name to be used to identify the overlay site.

6. ANOTHER: A yes/no switch which indicates that you have another site to overlay. Accept the default, YES to enter another site to overlay, enter NO if you have finished specifying overlay sites.

Once you have filled in the menu, enter NEXT. If the value of the parameter ANOTHER was YES, then the menu will be redisplayed so that you may enter another overlay site. If the value of ANOTHER was NO, you will continue to the next menu.

The next menu which will be displayed is the menu for overlay from a GEMS dataset.

MENU: Overlay from GEMS Dataset

ref	parmname	parameter description	value	index
1.	DATASET	GEMS overlay dataset name		
2.	METHOD	Location identification method		
3.	ZIP_NAME	Zip code variable name		
4.	LAT_NAME	Name latitude variable		
5.	LNG_NAME	Name longitude variable		
6.	OVR_NAME	Overlay identification variable	*	
7.	PLOTCHAR	Plot Character		
8.	PLTCOLOR	Plot Color	W	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS>

FIGURE 4-44. GEMS Dataset Overlay Menu

1. DATASET: The name of the GEMS dataset from which you would like to generate the overlay.

2. METHOD: The geographical location method to be used. Enter ZIPCODE to use zip codes as the location variable or LAT/LON to use latitude and longitude to specify the

overlay locations. If you had entered ZIPCODE, you will have to fill in a value for ZIP_NAME. If you had entered LAT/LON, you must provide values for LAT_NAME and LONG_NAME.

3. ZIP_NAME: The name of the variable which contains the zip codes. This must be a numeric variable.

4. LAT_NAME: The name of the variable which contains the latitude information.

5. LNG_NAME: The name of the variable which contains the longitude information.

6. OVR_NAME: The name of a character variable which contains descriptive text (e.g. site name) to label the overlay points with. If no variable is specified for this item the overlay points will not be represented in the legend.

7. PLOTCHAR: The plot character to be used to display the points on the map. You can use any character from the keyboard as a character or enter the following special plot symbols:

@C for a circle	@D for a diamond
@S for a square	@T for a triangle

8. PLTCOLOR: The color to use for this overlay. The colors to choose from are:

W - white	M - magenta
GY - green yellow	R - red
Y - yellow	BM - blue magenta
G - green	O - orange
RM - red magenta	B - blue
BC - blue cyan	GC - green cyan
C - cyan	DG - dark green
LG - light gray	

Depending on your terminal type you may have all or none of these colors available.

Enter NEXT to proceed when you have completed this menu. At this time, GEMS will read in the information from your overlay dataset. Depending on the size of your dataset, this could be a fairly long pause. Next, GEMS will prompt you if you want to enter another GEMS dataset. If you indicate YES, then you will see the same menu shown above. If you enter NO, then you will proceed to the next prompt where you will be given a chance to rescale your map.

Remember, you may not see all of the above prompts and menus, the path you follow will depend on the choices you have made in the main County Map menu. Next, GEMS will give you a chance to rescale your map. (This following prompt assumes you entered STATE as the method of location specification and elected to map Maryland).

The present scale factor is 1910000 or a 192 km radius.
Enter a new scale factor or study radius km
Default = 1910000
GEMS >

You can use this to increase the scope of the map beyond that which you have originally selected or scale the map down to plot less of the available data by increasing or decreasing the scale factor. You may enter the scale factor as a map scale or as a radius. To enter the scale as a radius enter the radius in kilometers followed by the letters km, e.g., "250 km". The space between the number and km is necessary. If you enter a value which is smaller than the given values, the following message is displayed:

You have requested a scale factor which will result in a map which contains less than the entire dataset. By default, the map will be expanded about the center of the screen; however, you may enter the latitude and longitude of a location around which you wish the expansion to occur.

Once you have answered this prompt, the map will be drawn. If you are on a Tektronix terminal, the map will be displayed for you at your terminal and then stored for later retrieval. If you are not on a terminal supporting graphics, then the new map dataset will be created which you can later review when you are at a terminal supporting graphics. An example of the output you might receive from the Draw County Map procedure is provided in Figure 4-45.

4.2.3 Hydrological Unit Map (HUMAP)

The Hydrological Unit Map procedure produces a map which has the outlines of the hydrological unit (HU) boundaries of any state in the U.S. A hydrologic unit is a code assigned by the U.S. Geologic Survey to delineate the drainage basin of a reach. Weather stations, HU identification values, data from a GEMS dataset, and other geographical areas of the U.S. may be overlaid on the hydrological unit map.

When you enter the HUMAP procedure, the prompt shown below appears.

Enter the state abbreviation:

Here you should enter the two-letter state abbreviation for the state you would like to map. GEMS then asks you:

Do you want to manually overlay a place? (Enter yes or no):

If you do not want to overlay a place manually (with you supplying the coordinates), then enter NO. If you have a place or places to overlay, enter YES. If you had entered YES, GEMS will next prompt you for the additional information necessary:

Data location overlay - Enter DONE to end.
The location methods are LAT/LONG and ZIPCODE
Please enter location method:

If you enter [illegible], then GEMS returns with the following prompt.

DEMO: NJ TANK SITES

1-BRUNSWICK
2-MONITOR
3-DOVER DE
4-WILMINGTON/GREATE DE
5-ATLANTIC CITY/NAF NJ
6-BELMAR/EVANS LAB NJ
7-NEWARK NJ
8-TETERBORO NJ
9-WRIGHTSTOWN/MCGUI NJ
10-NEW YORK/FT TOTTE NY
10-NEW YORK/LAGUARDI NY
10-N040000886
10-N040000882
10-N040000876D
10-N040000876C
10-N040000876B
10-N040000876A
10-N040000883
11-WHITE PL/WESTCHES NY
12-ALLENTOWN PA
13-PHILADELPHIA PA
14-WILKES-BARRE-SCRA PA
15-N040000890
15-N040000878
16-N040000892
16-N040000889
16-N040000885

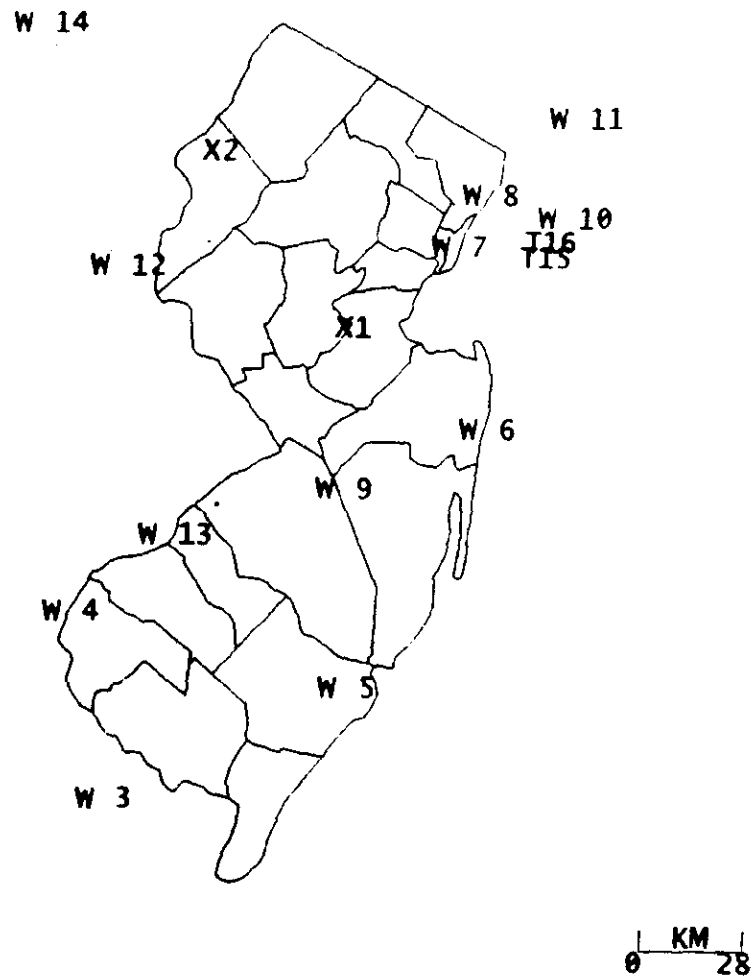

FIGURE 4-45. Example of a County Map

Enter the ZIPCODE:

Here you enter the zip code for the site you want to overlay. If you enter LAT/LONG, the following prompts are displayed:

Enter latitude:

Here you enter the latitude of the site to overlay. GEMS continues with:

Enter longitude:

Enter the longitude of the site to overlay. The next prompt is the same regardless of the method of location specification:

Enter overlay place name (30 characters max):

Here you enter the name to be used to identify the place which is being overlaid. GEMS will continue to ask for locations and place names until you enter **DONE**. GEMS will now prompt you for the rest of the options:

Do you want an overlay of weather stations? (Enter YES or NO):

If you would like to have the weather stations overlaid on the map, enter YES at the prompt. (Weather stations will be automatically retrieved by GEMS.) Enter NO, otherwise. The next prompt asks:

Do you want an overlay of HU ID values? (Enter YES or NO):

Reply with YES to overlay the Hydrologic Unit codes on the map; otherwise, enter NO. The HUMAP procedure can take information from a GEMS dataset for display on the map. The following prompt finds out if you want to use this feature:

Do you want an overlay from a GEMS dataset? (Enter YES or NO):

If you will not be overlaying from a GEMS dataset, enter NO. To overlay from a dataset enter YES. The HUMAP procedure will then ask you for the information necessary to do the overlay.

Enter the Dataset name:

Here enter the name of the GEMS dataset which will be used for the overlays. Once the dataset has been defined you will be asked to set up the variables from which the information is to be extracted:

Enter the location identifier:

The two methods of location identification are LAT/LONG and ZIPCODE. If you entered ZIPCODE, the procedure would prompt:

Enter the ZIPCODE variable name:

Here you enter the name of the variable in the dataset which contains the zip codes to use as overlay sites. If you had entered LAT/LONG, the zip code prompt would have been replaced by two prompts asking for the latitude variable and the longitude variable. Once you have identified the location variables you will be prompted for the name variable. This variable is used to generate the labels for the points being overlaid.

Enter the overlay NAME variable:

The variable used for the name must be a character variable. Next, you define the character to be used as the plot character for the map. This is the character which will be used for the overlay points.

Enter plot character:

For the plot character, any typeable character may be used. Once you have defined the plot character, you have finished the process for this dataset. You will be asked again if you

want to overlay from a GEMS dataset. If you want to use another dataset, enter YES; otherwise, enter NO to continue to the next step.

Enter the output map name:

Here you enter the name that the new map is to be stored under in your GEMS user library. You will then be asked:

Enter the tag field:

The tag field is a descriptive field of up to 60 characters which is used to describe the map being produced. Once the tag field has been defined, the following prompt appears.

Processing begins!!

GEMS now searches the Hydrological Unit boundaries file for the boundaries which are available for the state you have selected. You will then be prompted for the boundary to use, as in the example below.

The state contains 2 Datasets:
1 WISCONSIN N. HYDROLOGIC UNITS
STATE CODE = 55
2 WISCONSIN S. HYDROLOGIC UNITS
STATE CODE = 55
Enter a value 1 thru 2:

Once you have selected a value, the map heading is defaulted by the system. You are allowed to change it if you wish (for this example, we will assume you entered 2 above):

THE MAP HEADING IS:
WISCONSIN S. HYDROLOGIC UNITS
STATE CODE = 55
Enter a new heading or RETURN:

If you do want to change the heading for the map, enter the new heading at the prompt above; otherwise, you may simply enter a carriage return to continue to the next step:

The present Scale Factor is 22000000 or a 254 km radius.
Do you wish to change the Scale Factor (Enter YES or NO)?:

If you would like to change the scale factor for the map, enter a **YES** above. If you entered **YES**, the following prompt is displayed:

Enter the new Scale as a factor or a radius in km:

Enter your new scale at the above prompt. You may enter the scale factor as a map scale or as a radius. To enter the scale as a radius enter the radius in kilometers followed by the letters km, e.g., "250 km". The space between the number and km is necessary. If you enter a value which is smaller than the given values, the following message is displayed:

You have requested a scale factor which will result in a map which contains less than the entire dataset. By default, the map will be expanded about the center of the screen; however, you may enter the latitude and longitude of a location around which you wish the expansion to occur.

The following prompt then appears.

Do you want to enter a location (Enter YES or NO)?:

If you enter NO, the map will be centered in the display window. If you enter YES, the following prompts will be displayed to ask you where to center the map:

Enter a location method (ZIPCODE or LAT/LON):

Enter the location method you will be using. You will then be asked to enter the applicable parameters (the zip code or the latitude and longitude of the center point of the map).

The procedure will create the map. If you are on a Tektronix terminal, the map will be displayed at your terminal. The map will also be saved as a map dataset in your GEMS user library.

After your map has been created, GEMS will display the first prompt to begin creation of another map. An example of a map is provided in Figure 4-47. If you wish to end the procedure, enter EXIT and you will be returned to the GEOMAP menu.

4.2.4 The State/County Map Procedure

The State/County Map (STCOMAP) procedure produces choropleth maps of U.S. states and/or counties using data available from a GEMS dataset. The selected variable value must correspond to a geographic variable in your selected dataset, such as a state or county Federal Information Processing Standards (FIPS) code. If you select a state map, only state boundaries are drawn whereas, if you select a county map, state and county boundaries may be drawn. A state map may be drawn only if the state FIPS codes exist in the input dataset. For a county map, both the state FIPS code and the county FIPS code must be present. This procedure allows you the option of seeing all the counties or states in the area that you choose, even counties and states that have missing values which will be displayed in the map as "empty", i.e., they will not have patterns or colors assigned to them.

The State/County Map Procedure may be executed on all terminal types in the Terminal Type Specification menu, but you may only view your map on a Tektronix or Tektronix compatible terminal. The following parameter entry menu appears.

```
MENU: State/County Map

ref  parmname   parameter description                  value   index
1.  DATASET     Name of the input dataset
2.  OUTSET      Name of the output map dataset
3.  VARIABLE    Name of the variable to be mapped
4.  MAPOF       Map of: C(ounties) or S(tates)

Enter one or more combinations of: reference or parameter name and
value(s)[ref1value1,ref2value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-46. State/County Map Menu

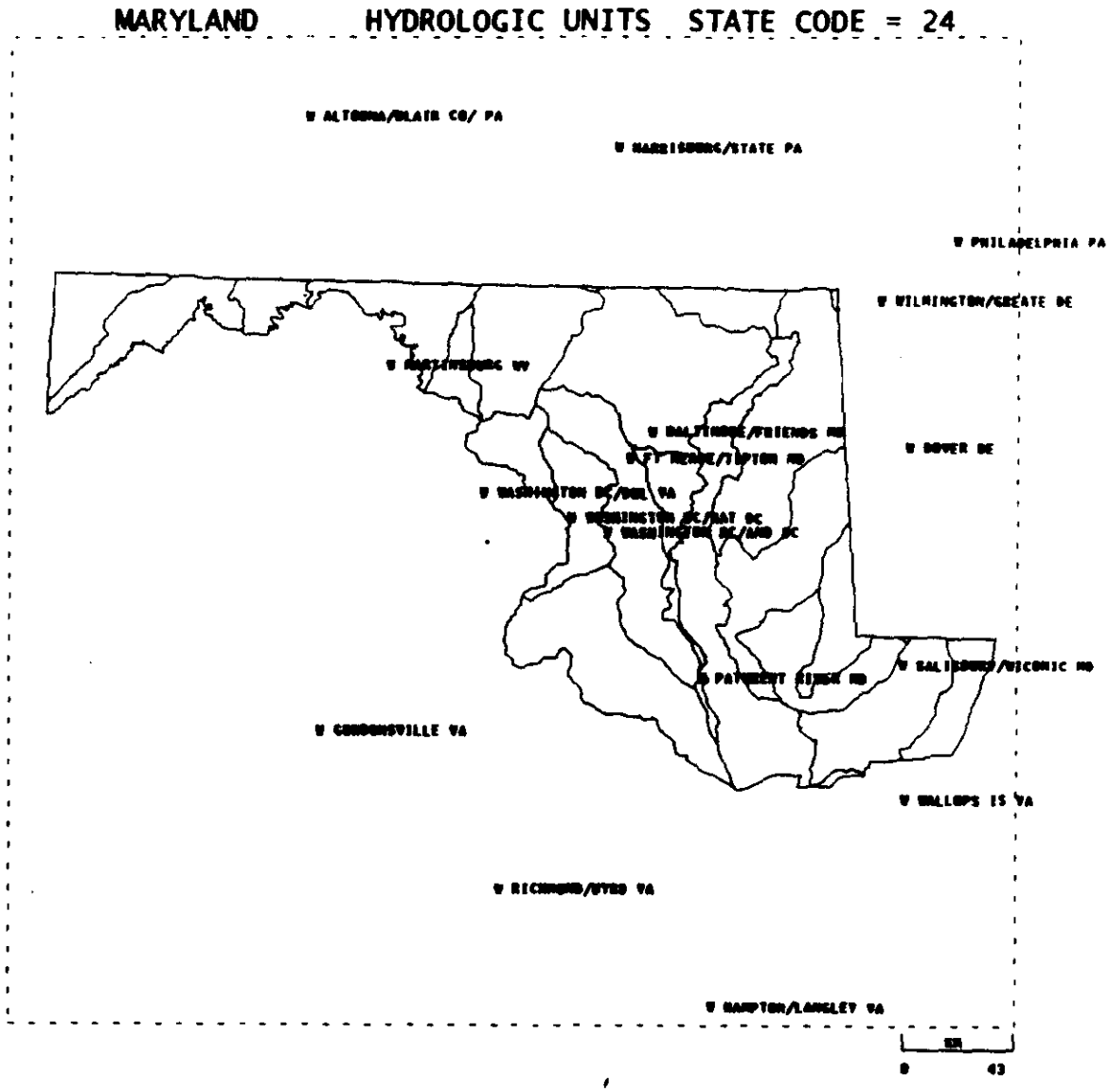

Figure 4-47. Example of Hydrological Unit Map

1. DATASET: The name of the dataset to be used as input to the procedure.

2. OUTSET: The name of the new map to be stored in your GEMS user library.

3. VARIABLE: The name of the variable to be used to produce the map. This variable must be present in the input dataset (DATASET).

4. MAPOF: This specifies the level of detail for the map to be created. Enter C to create a map for U.S. counties or an S to create a map with only the state boundaries.

Once the menu has been completed, the NEXT command allows you to proceed to the next step. The path you will follow is determined by your entry for Option 4. If you selected a county map it will be described in section 4.2.4.1. State maps are discussed in section 4.2.4.2.

4.2.4.1 County Map

This is the path which you will follow if you have selected the County map option from the first STCOMAP menu.

Enter the variable name containing state FIPS codes
GEMS >

Enter the variable name containing county FIPS codes
GEMS >

Enter valid variable names for the state and county FIPS codes. GEMS then validates your variable names and displays:

Dataset validation please wait...

Once the variable validation has completed, GEMS displays the Maps menu.

```
Menu: Choose One of the Following Maps:

1. Entire U.S. map                    (ENTMAP)
2. Continental U.S. map               (CONTMAP)
3. EPA regional map                   (EPAREG)
4. Census regional map                (CENREG)
5. Census division map                (CENDIV)
6. Selected States                    (STATMAP)

Enter an option number of a command: HELP,BACK,END,CLEAR,EXIT
GEMS >
```

FIGURE 4-48. Maps Menu

Enter your selected map type. Option 1, Entire U.S. map, draws a map of the entire U.S. including Alaska and Hawaii. Option 2, Continental U.S. map, draws a map only of the continental U.S. Option 3, EPA Regional map, allows the user to select one or more of ten EPA geographical regions for the map. Option 4, Census Regional map, allows the user to select one or more of four Census Regions for the map. Option 5, Census Division map, allows the user to select one or more of nine census divisions for the map. Option 6, Selected States, allows the user to select a map of one or more individual states.

Once you have selected your map type from the Maps menu, various prompts appear, depending upon your map type selection.

If you selected Option 1, Entire U.S. map, the following prompt appears.

Do you want to include outlines of the empty counties in the entire U.S.? (Y,N)
Default = N
GEMS >

If you wish to outline empty counties (counties that do not have data values assigned to them), enter **Y**. If you do not wish to outline the empty counties, enter **N**, or since it is the default value, a **carriage return**. If you entered NO, the following prompt appears.

Do you want state boundaries drawn in your map? (Y,N)
GEMS >

If you wish to outline the empty states (states that do not have data values assigned to them) in your map, enter YES. If not, enter NO.

If you selected Option 2, Continental U.S. map, the following prompt appears.

Do you want to include outlines of the empty counties in the continent? (Y,N)
Default = Y
GEMS >

If you wish to outline the empty counties (counties which do not have data values assigned to them) in your map, enter YES; otherwise, enter NO.

Do you want state boundaries drawn in your map? (Y,N)
GEMS >

If you wish to outline the empty states (states with no data) in your map, enter yes; otherwise, enter no.

If you selected Option 3, EPA Regional map, the EPA Region Selection menu appears.

Select one or more of the following EPA regions

Ref	Regions	States in the region
1.	I	CT, ME, MA, NH, RI, VT
2.	II	NJ, NY
3.	III	DE, MD, PA, VA, WV, DC
4.	IV	AL, FL, GA, KY, MS, NC, SC, TN
5.	V	IL, IN, OH, MI, MN, WI
6.	VI	AR, LA, NM, OK, TX
7.	VII	IA, KS, MO, NE
8.	VIII	CO, MY, ND, SD, UT, WY
9.	IX	AZ, CA, HI, NV
10.	X	AK, ID, OR, WA

Enter the reference number(s) of the regions wanted separated by blanks
GEMS >

FIGURE 4-49. EPA Region Selection Menu

Enter the EPA regions to be mapped. The states in each region are listed in the menu. EPA regions always include the entire state.

Once you have completed the EPA Regional Selection menu, the following prompt appears.

Do you want to include outlines of the empty counties in the region(s)? (Y,N)
Default = Y
GEMS >

If you wish to outline the empty counties in your map, enter YES; if not, enter NO. If you enter NO, this prompt appears.

Do you want state boundaries drawn in your map? (Y,N)

If you wish to outline the empty states (states with no data) in your map, enter yes; if not, enter no.

If you selected Option 4, Census Regional map, the parameter entry menu shown below appears.

Select one or more of the following Census regions

Ref	Regions	Divisions	States in the divisions
1.	NORTHEAST	New England	ME, NH, VT, MA, RI, CT
		Mid Atlantic	NY, NJ, PA
2.	NORTH CENTRAL	East north-central	OH, IN, IL, MI, WI
		West north-central	MN, IA, MO, ND, SD, NE, KS
3.	SOUTH	South Atlantic	DE, MD, DC, VA, WV, NC, SC, GA, FL
		West south-central	AR, LA, OK, TX
		East south-central	KY, TN, AL, MS
4.	WEST	Mountain	MT, ID, WY, CO, NM, AZ, UT, NV
		Pacific	WA, OR, CA, AK, HI

Enter the reference number(s) of the regions wanted separated by blanks
GEMS >

FIGURE 4-50. Census Regional Menu

Enter your Census region selection. The states which make up a Census region are listed with that region on the menu.

Once you have completed the Census Region Selection menu, the following prompt appears.

Do you want to include outlines of the empty counties in the region(s)? (Y,N)
Default = Y
GEMS

If you wish to outline the empty counties in your map, enter either [redacted] or a [redacted]. If you enter NO, this prompt appears.

Do you want state boundaries drawn in your map? (Y,N)
GEMS >

If you wish to outline the empty states in your map, enter YES; if not, enter NO.

If you selected Option 5, Census Division map, the Census Division Selection menu appears.

Select one or more of the following Census divisions:

Ref	Divisions	States in the division
1.	East north-central	OH, IN, IL, MI, WI
2.	Mid Atlantic	NY, NJ, PA
3.	Pacific	WA, OR, CA, AK, HI
4.	New England	ME, NH, VT, MA, RI, CT
5.	West north-central	MN, IA, MO, ND, SD, NE, KS
6.	South Atlantic	DE, MD, DC, VA, WV, NC, SC, GA, FL
7.	West south-central	AR, LA, OK, TX
8.	East south-central	KY, TN, AL, MS
9.	Mountain	MT, ID, WY, CO, NM, AZ, UT, NV

Enter the reference number(s) of the regions wanted separated by blanks
GEMS >

FIGURE 4-51. Census Division Selection Menu

Enter your Census division selection. The states which make up a Census division are listed along with the division in the menu. Once you have completed the Census Division Selection menu, the following prompt appears.

Do you want to include outline of the empty counties in the division(s)? (Y,N) Default = Y
GEMS >

If you wish to outline the empty counties in your map, enter YES. If you enter NO, this prompt appears.

Do you want state boundaries drawn in your map? (Y,N)

If you wish to outline the empty states in your map, enter YES. If not, enter NO.

If you selected Option 6, Selected States, the States Selection menu appears, displaying the following options.

Select one or more of the following states:

FIPS State	FIPS State	FIPS State	FIPS STATE
1. ALABAMA	18. INDIANA	32. NEVADA	47. TENNESSEE
2. ALASKA	19. IOWA	33. N. HAMPSHIRE	48. TEXAS
4. ARIZONA	20. KANSAS	34. NEW JERSEY	49. UTAH
5. ARKANSAS	21. KENTUCKY	35. NEW MEXICO	50. VERMONT
6. CALIFORNIA	22. LOUISIANA	36. NEW YORK	51. VIRGINIA
8. COLORADO	23. MAINE	37. N. CAROLINA	53. WASHINGTON
9.CONNECTICUT	24.MARYLAND	38.NORTH DAKOTA	54.W. VIRGINIA
10. DELAWARE	25. MASSACHUSETTS	39. OHIO	55. WISCONSIN
11. D.C.	26. MICHIGAN	40. OKLAHOMA	56. WYOMING
12. FLORIDA	27. MINNESOTA	41. OREGON	
13. GEORGIA	28. MISSISSIPPI	42. PENNSYLVANIA	
15. HAWAII	29. MISSOURI	44. RHODE ISLAND	
16. IDAHO	30. MONTANA	45. SOUTH CAROLINA	
17. ILLINOIS	31. NEBRASKA	46. SOUTH DAKOTA	

Enter the FIPS codes of the states wanted separated by blanks
GEMS >

FIGURE 4-52. State Selection Menu

Enter FIPS codes for the states you would like to map. The full list will be displayed with the menu to help you with your selection.

Once you have completed the States Selection menu, the following prompt appears.

Do you want to include outlines of the empty counties in the states selected? (Y,N)
Default = Y
GEMS >

Enter **Y** if you want outlines. Enter **N**, this prompt appears.

Do you want state boundaries drawn in your map? (Y,N)
GEMS >

Now you have completed the Maps Menu selection and prompts. Now, regardless of the type of map you selected you will continue on the Choropleth Map Options menu. Continue on to section 4.2.4.3 where the Choropleth Map Options menu is discussed.

4.2.4.2 State Map

If you select State Map for Option 4 of the State/County Map menu. When the prompt at the bottom of the menu appears, enter the NEXT command and the following prompt appears.

Enter the variable name containing state FIPS codes
GEMS >

Enter a valid variable name for the state FIPS code. The following prompt appears.

Dataset validation please wait...

If the dataset is valid, the Maps menu appears.

```
MENU: Choose One of the Following Maps:

1. Entire U.S. map
2. Continental U.S. map
3. EPA regional map
4. Census regional map
5. Census division map
6. Selected States

Enter an option number of a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 4-53. Maps Menu

Enter your selected map type. Option 1, Entire U.S. map, draws a map of the entire U.S. including Alaska and Hawaii. Option 2, Continental U.S. map, draws a map only of the continental U.S. Option 3, EPA Regional map, allows the user

to select one or more of ten EPA geographical regions for the map. Option 4, Census Regional map, allows the user to select one or more of nine census divisions for the map. Option 6, Selected States, allows the user to select a map of one or more individual states. Once you have selected your map type from the Maps menu, various prompts appear, depending upon your map type selection.

If you selected Option 1, Entire U.S. map, the following prompt appears.

Do you want to include outlines of the empty states in the the entire U.S.? (Y,N)
Default = Y
GEMS >

If you selected Option 2, Continental U.S. map, the following prompt appears.

Do you want to include outlines of the empty states in the continent? (Y,N)
Default = Y
GEMS >

If you selected Option 3, EPA Regional map, the EPA Regional Selection menu appears.

Select one or more of the following EPA regions:

Ref	Regions	States in the region
1.	I	CT, ME, MA, NH, RI, VT
2.	II	NJ, NY
3.	III	DE, MD, PA, VA, WV, DC
4.	IV	AL, FL, GA, KY, MS, NC, SC, TN
5.	V	IL, IN, OH, MI, MN, WI
6.	VI	AR, LA, NM, OK, TX
7.	VII	IA, KS, MO, NE
8.	VIII	CO, MY, ND, SD, UT, WY
9.	IX	AZ, CA, HI, NV
10.	X	AK, ID, OR, WA

Enter the reference number(s) of the regions wanted separated by blanks
GEMS >

FIGURE 4-54. EPA Regional Selection Menu

Enter your EPA region selection. Once you have completed the EPA Region Selection menu, the following prompt appears.

Do you want to include outlines of the empty states in the region(s)? (Y,N)
Default = Y
GEMS >

If you selected Option 4, Census Regional map, the Census Region Selection menu appears.

Select one or more of the following Census regions:

Ref	Regions	Divisions	States in the divisions
1.	NORTHEAST	New England	ME, NH, VT, NA, RI, CT
		Mid Atlantic	NY, NJ, PA
2.	NORTH CENTRAL	East north-central	OH, IN, IL, MI, WI
		West north-central	MN, IA, MO, ND, SD, NE, KS
3.	SOUTH	South Atlantic	DE, MD, DC, VA, WV, NC, SC, GA, FL
		West south-central	AR, LA, OK, TX
		East south-central	KY, TN, AL, MS
4.	WEST	Mountain	MT, ID, WY, CO, NM, AZ, UT, NV
		Pacific	WA, OR, CA, AK, HI

Enter the reference number(s) of the regions wanted separated by blanks
GEMS >

FIGURE 4-55. Census Region Selection Menu

Enter your Census region selection. Once you have completed the Census Region Selection menu, the following prompt appears.

Do you want to include outlines of the empty states in the region(s)? (Y,N)
Default = Y
GEMS >

If you selected Option 5, Census Division map, the Census Division Selection menu appears.

Select one or more of the following census divisions:

Ref	Divisions	States in the division
1.	East north-central	OH, IN, IL, MI, WI
2.	Mid Atlantic	NY, NJ, PA
3.	Pacific	WA, OR, CA, AK, HI
4.	New England	ME, NH, VT, MA, RI, CT
5.	West north-central	MN, IA, MO, ND, SD, NE, KS
6.	South Atlantic	DE, MD, DC, VA, WV, NC, SC, GA, FL
7.	West south-central	AR, LA, OK, TX
8.	East south-central	KY, TN, AL, MS
9.	Mountain	MT, ID, WY, CO, NM, AZ, UT, NV

Enter the reference number(s) of the Divisions wanted separated by blanks
GEMS>

FIGURE 4-56. Census Division Selection Menu

Enter your Census division selection. Once you have completed the Census Division Selection menu, the following prompt appears.

Do you want to include outlines of the empty states in the division(s)? (Y,N)
Default = Y
GEMS>

If you wish to outline the empty states in your map, answer YES. Otherwise, enter NO.

If you selected Option 6, Selected States, the States Selection menu appears, displaying the following options.

Select one or more of the following states:

FIPS State	FIPS State	FIPS State	FIPS STATE
1. ALABAMA	18. INDIANA	32. NEVADA	47. TENNESSEE
2. ALASKA	19. IOWA	33. N. HAMPSHIRE	48. TEXAS
4. ARIZONA	20. KANSAS	34. NEW JERSEY	49. UTAH
5. ARKANSAS	21. KENTUCKY	35. NEW MEXICO	50. VERMONT
6. CALIFORNIA	22. LOUISIANA	36. NEW YORK	51. VIRGINIA
8. COLORADO	23. MAINE	37. N. CAROLINA	53. WASHINGTON
9.CONNECTICUT	24.MARYLAND	38.NORTH DAKOTA	54.W. VIRGINIA
10. DELAWARE	25. MASSACHUSETTS	39. OHIO	55. WISCONSIN
11. D.C.	26. MICHIGAN	40. OKLAHOMA	56. WYOMING
12. FLORIDA	27. MINNESOTA	41. OREGON	
13. GEORGIA	28. MISSISSIPPI	42. PENNSYLVANIA	
15. HAWAII	29. MISSOURI	44. RHODE ISLAND	
16. IDAHO	30. MONTANA	45. SOUTH CAROLINA	
17. ILLINOIS	31. NEBRASKA	46. SOUTH DAKOTA	

Enter the FIPS codes of the states wanted separated by blanks
GEMS >

FIGURE 4-57. States Selection Menu

Enter the states' FIPS codes.

Once you have completed the States Selection menu, the following prompt appears.

Do you want to include outlines of the empty states in the states selected? (Y,N)
Default = Y
GEMS >

If you wish to outline the empty states in your map, answer YES. Otherwise, enter NO.

The Choropleth Map Options menu appears next, regardless of the type of map you selected. Continue with section 4.2.4.3 which discusses this menu.

4.2.4.3 Mapping Options

The following menu appears for both state and county map options. It allows you to specify the way your variable will be translated onto the chrolopleth map. This menu also allows you to add further identifying information such as titles and footnotes to the map.

```
MENU: Options (Choropleth Map)

ref parmname    parameter description                  value   index
1.  LLABEL      Legend label                            *
2.  COUTLNE     Color of Outline                        *
3.  CEMPTY      Color of empty areas                    *
4.  DISCRETE    Is variable discrete? (Y,N)             N
5.  LEGEND      Do you want a legend? (Y,N)             Y
6.  RANGES      Do you want to specify ranges? (Y,N)    N
7.  PATTERN     Do you want patterns? (Y,N)             N
8.  TITLE       Do you want titles? (Y,N)               N
9.  FTNOTE      Do you want footnotes? (Y,N)            N

Enter one or more combinations of: reference or parameter name and
value(s)[ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 4-58. Choropleth Map Options Menu

1. LLABEL: The label for the legend may be specified with this variable. The new label can be up to 15 characters. The default is the variable name.

2. COUTLNE: The color of the outline of the nonempty areas. The default is the same as the pattern color, which is selected in a later menu. You may choose your color from the following list: BLUE, MAGENTA, RED, CYAN, GREEN, WHITE, and YELLOW.

3. CEMPTY: The color of the outline for the empty areas. There is no color selected as a default. You may choose a color from the

following list: BLUE, MAGENTA, RED, CYAN, GREEN, WHITE, and YELLOW.

4. DISCRETE: A yes/no switch which allows you to have a different pattern for each distinct value of the variable. If the variable you are using has no more than 20 different values, then you may choose to use the discrete option. (If there are more than 20 values, you may specify ranges.)

5. LEGEND: A yes/no switch which indicates whether you wish to generate a legend for the range of values assigned to a pattern. To generate a legend for the range of values assigned to a pattern, enter Y for this option. For no legend, enter N. The default is to generate a legend. The legend will be created by SAS automatically below the map.

6. RANGES: A yes/no switch which allows you to specify the ranges of values for the variable which is being mapped. You may specify up to 12 ranges of values. Each range of values will be shaded or colored differently. If you are generating a legend for the map, the ranges and the patterns for the ranges will be in the legend.

7. PATTERN: A yes/no switch to change default patterns. The procedure uses 12 default patterns. If you would like to change these patterns or add more for this run, enter Y for this option and a menu will appear asking for those patterns. PATTERN statements must be specified for each level of the variable. The number of levels which will be used is:

- The number of discrete values the variable has, if you selected DISCRETE.

- The number of ranges which you have specified, if you elected to specify ranges.

- If neither of these options were selected, then the procedure uses the formula: 1 + 3.3LOG(N) where N is the number of state or county FIPS values in your dataset.

8. TITLE: A yes/no switch which allows you to specify the title for the map. You may specify up to 10 lines to be printed at the top of the map. Should you wish to do this, enter Y. The default is not to print titles. If you select this option you will later have the option to specify color, type font, type size and position for the text.

9. FTNOTE: A yes/no switch which allows you to specify the footnote for the map. You may specify up to 10 lines to be printed at the bottom of the map. Should you wish to do this enter Y. The default is not to print footnotes. If you select this option you will later have the option to specify color, type font, type size and position for the text.

Once the Choropleth Map Options menu is completed, enter the [redacted] command. For each choropleth map option you selected from the Choropleth Map Options menu, an Options menu or prompt is displayed. For example, if you entered "6 Y", "7 Y", "8 Y", and "9 Y", the following sequence of prompts and menus appear:

Enter the number of ranges
GEMS >

You may specify from two to twelve distinct divisions or "ranges" for your variable value. Once you have entered the number of ranges, the following prompts appear:

Enter the upper limit for the 1st range
GEMS >

Enter the label for the 1st range
Default = MIN
GEMS >

The lower limit for the first range is the same as the lowest value for the variable. Therefore, you will not be prompted to enter the lower limit for the range. All subsequent lower ranges are computed by GEMS and are based on the previous upper limit entered. You may enter exactly the upper value and label for each range. GEMS continues to prompt for range and label for the number of ranges you specified.

If you specified Y for PATTERN on the Choropleth Map Options menu, the Map Pattern Options menu appears.

MENU: Map Pattern Options

ref	parmname	parameter description	value	index
1.	NUM	Pattern number	1	
2.	TYPE	Pattern type (X,L,S,E)		
3.	DENSITY	Line density, specify when TYPE = X,L	1	
4.	ANGLE	Line angle, needed when TYPE = X,L	0	
5.	COLOR	Pattern Color	BLUE	
6.	REP	Number of Repetitions	1	
7.	MORE	Do you want more patterns? (Y,N)	N	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command:HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-59. Map Pattern Options Menu

1. NUM: The number of this pattern. The pattern number corresponds to the first, second... etc. ranges or discrete values for your map. There have been twelve

(12) default patterns already defined. You may change any of these or add additional patterns to the list. To alter a pattern, enter that pattern number as NUM. If you enter more patterns than there are ranges or discrete values for your map, the additional patterns will not be used.

2. TYPE: The type of pattern, available types are:

E - the empty pattern
S - for a solid pattern
L - for a parallel line pattern
X - for a cross-hatched line pattern

Note that the density and angle options in the menu have no effect on either the empty or solid patterns.

3. DENSITY: The density of the lines which will be used to make up the pattern. You should specify a value between 1 and 5 where 1 is the lightest shading and 5 is the darkest.

Density only applies to the X (cross hatch) and the L (parallel lines) pattern types.

4. ANGLE: The angle of the pattern lines. You may enter values from 0 to 360 degrees.

The Angle parameter only applies to the X (cross hatch) and the L (parallel lines) pattern types.

5. COLOR: The color for your patterns. The available colors are: BLUE, RED, GREEN, YELLOW, MAGENTA, CYAN, and WHITE.

6. REP: The number of times the pattern will be repeated for consecutive ranges or discrete values. This allows you to use a pattern to define a number of ranges or discrete values without having to specify it each time. Please note that if you do have a pattern repeated for consecutive ranges, you will not be able to diffrentiate between levels.

7. MORE: A yes/no switch which allows you to specify/define more patterns. If the value of this pattern is N, you will continue to the next step, whereas if the value is Y the procedure will return to the Map Pattern Options menu so that you may specify additional patterns.

Once this menu is completed, enter NEXT. If you entered Y for the TITLE option on the choropleth menu, the following menu is displayed. The map title can be made up of a number of lines of text which you can specify on this screen. Each line must be specified separately on its own screen.

MENU: Map Title

ref	parmname	parameter description	value	index
1.	NUM	Line number of the title	4	
2.	TEXT	Content of the title	*	
3.	COLOR	Color of the title	GREEN	
4.	FONT	Font type of the title	*	
5.	JUST	Justification of the title	*	
6.	HEIGHT	Height of title characters	1	
7.	ANGLE	Angle of the title line (deg)	0	
8.	ROTATE	Angle of title characters	0	
9.	MOVE	Coordinates of title movement	0	
10.	MORE	More titles? (Y,N)	N	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command:HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-60. Map Title Menu

1. NUM: A number from 1 to 10 which specifies the line on which to put this title. The title can be made up of up to ten lines of text.

2. TEXT: The text for this title line. Each line can be up to 64 characters of text.

3. COLOR: The color for the title. The available colors are: RED, GREEN, BLUE, YELLOW, MAGENTA, CYAN, and WHITE.

4. FONT: The font for the title. The available fonts are: SIMPLEX, DUPLEX, TRIPLEX, COMPLEX, OLDENG, SCRIPT, ITALIC, and GITALIC.

5. JUST: The justification scheme. You may choose one of the following justification schemes for the title:

C or CENTER	for a centered title(default)
R or RIGHT	for a right justified title
L or LEFT	for a left justified title

6. HEIGHT: The size of each character of the title. You must specify a number between 0.5 and 4.0. The default character size is 1.

7. ANGLE: The angle of rotation of the title line. You specify the angle by entering a number between -90 and 90 degrees. The default angle is 0 degrees which indicates a horizontal title line. A negative value will rotate the title line clockwise and a positive angle will rotate the title counter-clockwise.

8. ROTATE: The angle of rotation of each character in the title. You may choose an angle between 0 and 360 degrees. The default angle is 0 degrees which indicates an upright position.

9. MOVE: Allows you to move the title. You may move the title by the specified coordinates. The coordinates are given by character cells, rows and columns on the display.

10. MORE: A yes/no switch to add additional title lines. In order to add additional title lines to the map by entering Y for this parameter. Entering N ends title entry when you complete this menu.

When the Map Title menu is completed, enter NEXT in order to continue. If you selected the map footnote option, the following menu will be displayed.

MENU: Map Footnote

ref	parmname	parameter description	value	index
1.	NUM	Line number of footnote	1	
2.	TEXT	Content of the footnote	*	
3.	COLOR	Color of the footnote	*	
4.	FONT	Font type of the footnote	*	
5.	HEIGHT	Height of footnote characters	1	
6.	JUST	Justification of the footnote	*	
7.	ANGLE	Angle of footnote line (deg)	0	
8.	ROTATE	Angle of footnote characters (deg)	0	
9.	MOVE	Coordinates of footnote movement	0	(2)
10.	MORE	More footnotes? (Y,N)	N	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 4-61. Map Footnote Menu

1. NUM: A number from 1 to 10 which specifies the line on which to put this title. The title can be made up of up to ten lines of text.

2. TEXT: The text for this title line. Each line can be up to 64 characters of text.

3. COLOR: The color for the title. The available colors are: RED, GREEN, BLUE, YELLOW, MAGENTA, CYAN, and WHITE.

4. FONT: A font for the title. The available fonts are: SIMPLEX, DUPLEX, TRIPLEX, COMPLEX, OLDENG, SCRIPT, ITALIC, and GITALIC.

5. JUST: The justification scheme. You may choose one of the following justification schemes for the title:

C or CENTER	for a centered title
R or RIGHT	for a right justified title
L or LEFT	for a left justified title

The default is to have the title centered.

6. HEIGHT: The size of each character of the title. You must specify a number between 0.5 and 4.0. The default character size is 1.

7. ANGLE: The angle of rotation of the title line. You specify the angle by entering a number between -90 and 90 degrees. The default angle is 0 degrees which indicates a horizontal title line. A negative value will rotate the title line clockwise and a positive angle will rotate the title counter-clockwise.

8. ROTATE: The angle of rotation of each character in the title. You may enter an angle between 0 and 360 degrees. The default angle is 0 degrees which indicates an upright position.

9. MOVE: Specifies whether the title is to be moved. You may move the title by the specified coordinates. The coordinates are given by character cells, rows and columns on the display.

10. MORE: A yes/no switch which specifies whether additional title lines are needed. You may add additional title lines to the map by entering Y for this parameter. Entering N ends title entry.

Once you have finished all of the menus which you have selected, GEMS will ask you how to run the program.

Enter program execution mode: B (batch) or I (interactive)
GEMS >

Batch mode execution allows you to return to GEMS and continue working while your map is being created. When the map is complete, GEMS displays a message wherever you are in that procedure stating that the creation of the map is finished. Interactive mode execution completes the map while you wait.

When the map is completed in the interactive mode, it is composed on the screen while you wait. To exit from viewing your county map, press the Carriage return, and you will be returned to the GEOMAP menu. In either case, the map will be stored for later retrieval. An example of a State/County map is provided in Figure 4-61.

CHOROPLETH MAP OF COUNTIES

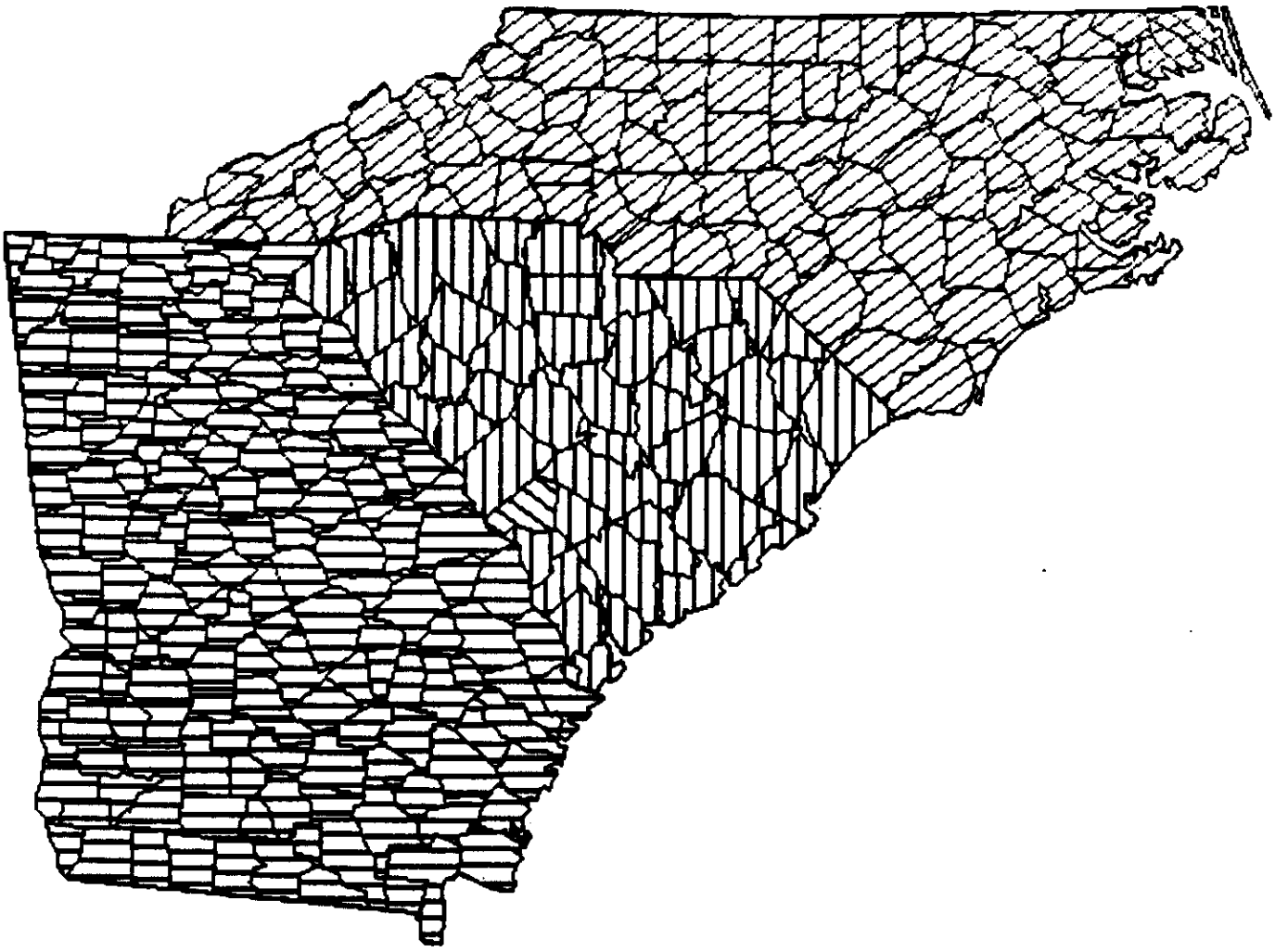

Figure 4-62. Example of State/County Map Output

Chapter 5 - GEMS File Management Facility

Many procedures in GEMS create or use data files. The GEMS File Management facility was created to manage these files. It supports the creation, modification, and handling of GEMS files as well as their storage and use. Files which are created by GEMS functions are stored under the GEMS File Management facility. They may be stored as DATASETS, GRAPHS, TABLES, MAPS, and REPORTS. A description of each follows.

DATASETS are files which have been created by data entry, importing a VAX file into GEMS, or through the use of one of the other functions under GEMS which creates datasets. There are also several on-line GEMS datasets which may be accessed using the File Management facility. (See Chapter 9 for more information on these datasets.) Datasets hold information at the variable level and can be manipulated through use of several GEMS functions including models and mapping.

GRAPHS are files which have been created by the graphics procedures or by the more specialized graphics functions available under certain GEMS models.

TABLES are tabular listings of data which have been written out by one of the GEMS functions, including Statistics (Chapter 6).

MAPS are created by the GEMS mapping functions available under the Geodata Handling procedure.

REPORTS are model output files which have been generated by the models installed in GEMS.

Figure 5-1 is a reproduction of the File Management menu. The first three capabilities are self-explanatory. You can use the GEMS File Management OUTPUT procedure to display any of the above data types. The EXPORT and IMPORT capabilities allow the transfer of files to and from the VAX level.

To determine what files are contained in your GEMS user dataset library you can use the LIST command at any menu screen in GEMS. Simply follow the command with the data type which you would like to look up, for example, LIST DATASET. Remember that there are several components of GEMS which have their own user interface, the EXAMS model for example, and the normal GEMS commands and features are not available within these components. When you exit from such a procedure, you will be returned to the GEMS interface where you will have access to all GEMS facilities again.

```
MENU: File Management

1. Create Datasets                                  (CREATE)
2. Modify Datasets                                  (MODIFY)
3. Delete Datasets/Graphs/Tables/Maps/Reports       (DELETE)
4. Output Datasets/Graphs/Tables/Maps/Reports       (OUTPUT)
5. Export Datasets/Graphs/Tables/Maps/Reports       (EXPORT)
6. Import Datasets/Graphs/Tables/Maps/Reports       (IMPORT)
7. Change Status of Datasets/Graphs...etc.          (STATUS)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 5-1. File Management Menu

Each of the items in the menu above will be described in sections that follow.

5.1 Create Datasets (CREATE)

The CREATE option gives you access to a number of functions which can be used to create a completely new dataset or to perform an operation on an existing dataset (such as SORT) and create a new dataset in the process. When you select this option, the menu in Figure 5-2 appears. Each of the menu options are discussed in this section.

You should remember that whenever you create a new dataset (by whatever process), it is created as a temporary dataset. Temporary datasets are automatically deleted at the end of your GEMS session unless you save them. When you exit GEMS, you will be given a list of all temporary datasets which have been created and asked if you would like to save any of them. You can save any or all datasets as needed.

```
MENU: Create Datasets

1. Create a dataset by data entry          (ENTRY)
2. Use the VT100 dataset editor            (VTCREATE)
3. Sort a dataset                          (SORT)
4. Merge two datasets                      (MERGE)
5. Concatenate two datasets                (CONCAT)
6. Functional data extraction              (FUNEXT)
7. Extract data from a dataset             (EXTRACT)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 5-2. Create Datasets Menu

5.1.1 Create Datasets by Data Entry

This option allows you to create a new GEMS dataset in your user dataset library. This new dataset can contain up to a total of twenty variables. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Create a Dataset by Data Entry
ref   parmname     parameter description          value   index
1.    DATASET      Name of the output dataset
2.    NUMRECS      Number of data records
3.    TAG          Tag field of the output dataset   *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 5-3. Create Dataset by Data Entry Menu

1. DATASET : The name of the new dataset which you are creating. This name must be unique in the system (i.e., neither a user dataset nor a system dataset can exist which already has this name). If you enter a name used by a system dataset, it will not be allowed. If the name is already in use in your user dataset library, you will be given the option to delete the existing dataset (supersede it) and replace it with the new dataset.

2. NUMRECS : The number of records in the dataset. All variables must have the same number of records when they exist in the same dataset.

3. TAG : A descriptive field to help you in identifying the contents of the dataset in the future. It can be up to 60 characters long.

After completing the menu, enter NEXT to proceed. You will recieve the following prompt:

Enter the name of the first variable:
GEMS >

Name the variable. The variable names must be unique within a dataset. You will then see:

Enter variable type: I (integer), R (real), C (character string)
GEMS >

GEMS understands three different variable types. They are:

1. INTEGER : Whole numbers only
2. REAL : Floating point numbers (decimal numbers)
3. CHARACTER STRINGS: Simply treated as a group with no internal processing or checking performed.

At this prompt you must pick the one which represents the values which you are now entering. This prompt will be followed by:

Enter {NUMRECS} values
{VARNAME} N
GEMS >

In the above example, GEMS will substitute the proper number of values for {NUMRECS} and the variable name which you have entered for {VARNAME} when this prompt is displayed. You must enter the values for the variable which you are currently defining. You will see successive prompts until you have entered the number of values which you defined in the first menu. If you are entering numeric values, you should know that a -999999 is interpreted by GEMS as a missing value. You can use this, if necessary, as a "place-holder".

Once you have entered all values for this variable, you are done with it and the program will continue to the next variable. The prompt **Enter the name of the next variable:** will appear and the loop starts over again. To end this process, simply enter a carriage return. Upon completion, a message will be displayed: **Dataset {NAME} successfully created.**

Remember that any new datasets which are created during a session are temporary and will be deleted at the end of the session if you do not save them.

5.1.2 Use VT100 Dataset Editor (VTCREATE)

This procedure provides another technique of data entry which is very useful for you if you have access to VT100 terminals. VTCREATE enables you to enter your data using a full screen data editor which also allows you to move around within the file to correct any errors you might have made during the entry process. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: VT100 Dataset Editor
ref   parmname     parameter description          value   index
1.    DATASET      Name of the output dataset
2.    TAG          Tag field of the output dataset   *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 5-4. VT100 Dataset Editor Menu

1. DATASET : The name of the new dataset which you want to create using the data editor.

2. TAG : The description field which will be associated with the new dataset. It can be up to 60 characters long.

After completing this menu, enter NEXT to proceed to the next screen. The parameters in this menu are also defined below.

```
MENU: Define Variable #1
ref   parmname     parameter description          value   index
1.    VARIABLE     Name of the output dataset
2.    DATATYPE     Data type of the variable (I,R,C)
3.    WIDTH        Width of the value field
4.    MORE         More variables? (Y,N)            Y

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 5-5. Define Variable #1 Menu

1. VARIABLE : The name for the first variable in the file.

2. DATATYPE : The type of variable. Valid variable types are (I)nteger, (R)eal and (C)haracter.

3. WIDTH : The column width allocated for input of the data. This will be taken as the data width for character variables in the new file; however, for numeric variables it is only used for data entry purposes.

4. MORE : A flag to determine if you have more variables to be set up for this file. If you do have more variables, then leave the default YES. If you have finished with your variable definitions, enter N for this item. If you answer YES, you will continue to get additional screens just like this one to fill out. Once you have finished and answer N, you will continue to the next step.

After completing this screen, enter NEXT. The program will continue to the next screen if you are defining more variables. Otherwise, you will enter the data editor. The data entry screen will look something like this:

```
GEMS DATA EDITOR    Dataset: {NAME}        Fields: 1  Records: 1
*************************************************************
REC #  {VARNAME}
1
*************************************************************

Position data field with arrow keys and enter data

KEYPAD: 1 HELP 2 PGUP 3 PGDOWN 4 PGLEFT 5 PGRIGHT 6 TOP
7 BOTTOM 8 DONE 9 QUIT
```

FIGURE 5-6. GEMS Data Editor Menu

The keypad refer to the keys on the numerical keypad on the right of the VT100 keyboard. The field which you are currently editing will be highlighted on your screen. After entering your data for the field, use a *carriage return* to move onto the next field. Remember you can always use the arrow keys to move back to a field which you have entered previously to correct any mistakes you may have made.

The keypad key definitions are:

KEY	DEFINITION
1	Display help information on how to use the editor.
2	Move back one page.
3	Move forward one page.
4	Move to the left one page.
5	Move to the right one page.
6	Move to the top of the dataset.
7	Move to the last page of the dataset.
8	Done with data editing. Leave the editor and create the file using the entered data.
9	Quit editing. Leave the editor, but do not create a file. (Abort the procedure)

As you saw in Figure 5-6, these definitions are displayed across the bottom of the entry screens.

When you finish your data entry, enter 8 from the keypad to signal that you have finished. You will then leave the editor and the new dataset will be created containing the data you have just entered. You will return to the main CREATE DATASETS menu. If you abort the procedure by entering 9, no dataset will be created and you will be returned to the main CREATE DATASETS menu.

5.1.3 Sort a Dataset (SORT)

The SORT procedure allows you to create a new dataset by sorting key variables within an existing dataset. The dataset can be sorted in either ascending or descending order. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

MENU: Sort a Dataset

ref	parmname	parameter description	value	index
1.	INSET	Name of the input dataset		
2.	KEYS	Names of key variables		(1)
3.	ORDER	Sorting order: A or D		(1)
4.	OUTSET	Name of the output dataset		
5.	TAG	Tag field of the output dataset		*

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-7. Sort a Dataset Menu

1. INSET : The name of the dataset which you want to sort. The output dataset will contain the same information as this dataset, but in sorted order.

2. KEYS : The variables which will be used to sort the dataset. The first variable will be the primary key (i.e., the order the dataset will be sorted in). Any additional variables will be used to sort the dataset within the

order already defined by the variables which were specified before it.

3. ORDER : The order in which the dataset is to be sorted. Enter A for ascending order or D for descending order.

4. OUTSET : The name of the new sorted dataset which is being created.

5. TAG : A description field which may be up to 60 characters long to help you remember what is in this dataset in the future.

Fill in the menu with your choices; enter NEXT to proceed to the next step where you will see:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

If you are sorting a large dataset, batch execution is recommended. Enter your choice and the sort will start at your terminal or you will see a message informing you that the batch job has been submitted. When the sort has been completed, you will be returned to the CREATE DATASETS menu.

5.1.4 Merge Two Datasets (MERGE)

The MERGE function allows you to create a new dataset from two existing datasets by attaching the records from the second dataset to those of the first. Optionally you may specify a keyed merge in which case the datasets are first sorted by the specified key variables and then merged.

Any two datasets may be merged to produce a new dataset. The amount of space available on your disk governs the size of the datasets which you may create, but we recommend that you not merge datasets that are very large. The parameter entry screen is shown below, and a discussion of each parameter follows.

MENU: Merge Two Datasets

ref	parmname	parameter description	value	index
1.	INSET1	Name of the first input dataset		
2.	INSET2	Name of the second input dataset		
3.	KEYMERGE	Key merge? (YES or NO)		
4.	KEYVAR1	Key variable of the first dataset		*
5.	KEYVAR2	Key variable of the second dataset		*
6.	OUTSET	Name of the output dataset		
7.	TAG	Tag field of the output dataset		*

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-8. Merge Two Datasets Menu

1. INSET1 : The name of the first input dataset to be used in the merge.

2. INSET2 : The name of the second input dataset to be used in the merge.

3. KEYMERGE : A YES here indicates a keyed merge, NO means that the records will be attached in the current order.

4. KEYVAR1 : The name of the variable to be used as the key for the first dataset (i.e., the variable by which to sort the records in the first dataset). Required only if performing a keyed merge.

5. KEYVAR2 : The name of the variable to be used as the key for the second dataset (i.e., the variable by which to sort the second dataset). Required only if performing a keyed merge.

6. OUTSET : The name for the new dataset to be created by the merge.

7. TAG : A description field which may be up to 60 characters long.

Complete the menu and enter NEXT to proceed to the next step where you will see:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

Batch execution is recommended unless these are very small datasets. Once you have made your choice, the procedure will be executed. If you have selected batch mode, the system will display a message informing you of the creation of the batch job. You will then be returned to the CREATE DATASETS menu. (If you selected interactive mode, you will be unable to proceed until the procedure is completed.)

5.1.5 Concatenate Two Datasets (CONCAT)

This procedure creates a new dataset by adding the records from the second dataset at the end of the records from the first dataset.

The difference between the CONCAT function and the MERGE function is that the CONCAT function requires two datasets with the same organization. This function simply adds records from one dataset to the end of the other dataset. The MERGE function, on the other hand, can merge two dissimilar datasets together with no variables in common (unless you are doing a keyed merge, in which case there must be at least one variable in common).

Before you begin concatenation, there are a couple of important things to remember:

- Both files you will use in the concatenation must be of identical format and have identical variables.
- The sequence of the two files must also be the same. If the files have the same format and the same variables, but in a different order, the process will not work!

Before you begin concatenation, verify the format and sequence of both datasets. (You can do this using the Output option of File Management, described in section 5.4.1.) The two source files used are unchanged. The parameter entry menu for this option is shown below, and a discussion of each paramenter follows.

MENU: Concatenate Two Datasets

ref	parmname	parameter description	value	index
1.	INSET1	Name of the first input dataset		
2.	INSET2	Name of the second input dataset		
3.	OUTSET	Name of the output dataset		
4.	TAG	Tag field of the output dataset		*

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-9. Concatenate Two Datasets Menu

1. INSET1 : The first dataset which is to be used in the concatenate operation.

2. INSET2 : The name of the second dataset which is to be used in the concatenate operation.

3. OUTSET : The name of the new dataset which will be created which holds the records from the two input datasets.

4. TAG : A description field which may be up to 60 characters long.

Complete the menu and enter NEXT to proceed to the next step where you will see the prompt:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

Batch execution is recommended. Once you have made your choice, the procedure will be executed. If you have selected batch mode, the system will display a message informing you of the creation of the batch job. You will be returned to the CREATE DATASETS Menu.

5.1.6 Data Transformation (TRANSFORM)

The Data Transformation procedure allows you to create a subset of an existing dataset by extracting specific variables and by creation of new variables through arithmetic operations. The calculation of a new variable is accomplished by specifying a new variable name followed by an equal sign and the equation to be used. Items to the right of the equal sign must be variables already defined in the inset dataset or constants connected by a mathematic operator or function. Acceptable mathematical operators and functions include: plus (+), minus (-), multiplication (*), division (/), exponentiation (**), concatenation (), LOG10 (base 10), LOG and EXP. The HELP command offers examples of equations.

You can also simply extract specific variables from a dataset. At the "equation" parameter, enter the new variable name followed by the equal sign, with the name of the input variable after the equal sign. For more than one variable, enter only the reference number (3) for the "equation" parameter. GEMS will then prompt you for a series of new/input variable pairs. The parameter entry menu for this option is shown below.

MENU: Data Transformation

ref	parmname	parameter description	value	index
1.	INSET	Name of the input dataset		
2.	OUTSET	Name of the output dataset		
3.	EQUATION	Equations for the output variables		(1)
4.	TAG	Tag field of the output dataset	*	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-10. Data Transformation Menu

1. INSET : The name of the input dataset whose variables you want to be transformed.

2. OUTSET : The name of the new output dataset which will be created by this procedure.

3. EQUATION : The equation(s) to be used to perform the extractions or transformations to be done during the extract operation. See the rules above for how to format an equation for extract. If you are entering more than one equation, enter only the "ref" number and GEMS will revert to prompt mode and prompt you for that parameter. Be sure to check over your equations before proceeding.

4. TAG : A description field of up to 60 characters.

After completing the menu with your choices, enter *NEXT* to proceed to the next step where you will see:

Enter program execution mode: B (batch) or I (Interactive)
GEMS >

Batch execution is recommended unless these are very small datasets. Once you have made your choice, the procedure will be executed. If you have selected batch mode, the system will display a message informing you of the creation of the batch job. You will be returned to the CREATE DATASETS Menu. (If you had selected interactive mode, you will be unable to proceed until the procedure is completed.)

5.1.7 Extract Data From a Dataset (EXTRACT)

This procedure allows you to create subsets of existing GEMS datasets from larger existing datasets. You may set the criteria, such as the value(s) or range of values for selected variables to filter out undesired data. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

MENU: Extract Data from a Dataset

ref	parmname	parameter description	value	index
1.	INSET	Name of the input dataset		
2.	VARSRCH	Names of variables to search on	*	(1)
3.	VARVALUE	Values of variables to search on	*	(1)
4.	LOGIC	Logical option (AND, OR)	AND	
5.	OUTVAR	Names of the output variables	*	(1)
6.	OUTSET	Name of the output dataset		
7.	TAG	Tag field of the output dataset	*	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-11. Extract Data from a Dataset Menu

1. INSET : The name of the dataset from which you wish to make a subset.

2. VARSRCH : The search variables to be used in making the extracted subset. You may choose any variable present in the input data set.

3. VARVALUE : The search criteria to be applied to the search variable. You may specify a specific value or a range of values in the following format.

INPUT FORMAT	EXAMPLE
value	1
minimum:	0:
maximum:	:100
min:max	0:100

You may also specify a sub-string for a character variable to be searched. For example, if the state/county FIPS code is stored as a 5-digit character string you could enter 19 to retrieve all values for state FIPS code 19.

4. LOGIC : If you have selected more than one search variable, you must specify the logic to be applied. If you select AND, a record must meet all specified search criteria in order to be extracted; should you select OR, it must only meet one of the specified criteria.

5. OUTVAR : The name(s) of the variable(s) which should be included in the new output dataset. It should be noted that the search variables are not automatically selected. If you want them in the new dataset, you must list them here again.

6. OUTSET : The name of the output dataset which will be created with the extracted variables.

7. TAG : A description field of up to 60 characters to remind you of what is in the output dataset.

Once you have filled the menu with your choices enter NEXT to proceed to the next step where you will see the following prompt:

Enter program execution mode: B (batch) or I (interactive)
GEMS >

Batch execution is recommended unless these are very small data sets. Once you have made your choice, the procedure will be executed. If you have selected batch mode, the system will display a message informing you of the creation of the batch job. You will be returned to the CREATE DATASETS Menu. (If you had selected interactive mode, you will be unable to proceed until the procedure is completed.)

5.2 Modify Datasets (MODIFY)

The MODIFY procedures allow you to edit a dataset to change values in it, add records, delete records or to modify attributes such as the name of the dataset or the names of the variables contained in the dataset.

You may only edit or alter attributes for the datasets contained in your user dataset library. You may not alter system-installed datasets.

```
MENU: Modify Datasets

1. Edit a Dataset                          (EDIT)
2. Use the VT100 Data Editor               (VTMODIFY)
3. Add Records to a Dataset                (ADDREC)
4. Delete Records from a Dataset           (DELREC)
5. Modify Dataset Attributes               (ATTRIBUTE)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 5-12. Modify Datasets Menu

Just enter a number between 1 and 5 to modify your dataset.

5.2.1 Edit a Dataset (EDIT)

This procedure allows you to modify the values in an existing GEMS dataset. When you invoke this procedure, you will be first asked some questions to determine which dataset you would like to edit.

Enter the name of the input dataset
GEMS >

Once the dataset has been selected, you must then specify the variable which you would like to edit:

Enter the name of the variable to be edited

Now you must enter the record number to be changed and the new value to be assigned to that record number:

Enter a record number
GEMS >

Enter a new value
GEMS >

To end editing, enter a *carriage return* at the **Enter a record number** prompt. You will be returned to the **Enter the name of the variable to be edited** prompt. You may then choose another variable or enter a *carriage return* to end the procedure.

5.2.2 Use the VT100 data editor (VTMODIFY)

This procedure allows you to use the VT100 data editor to modify an existing dataset's values. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

MENU: Modify a Dataset by VT100 Data Editor

ref	parmname	parameter description	value	index
1.	DATASET	Name of the input dataset		
2.	VARIABLS	Names of the variables	*	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-13. Modify a Dataset by VT100 Data Editor Menu

1. DATASET : The name of the dataset which you wish to modify.

2. VARIABLS : The variables which you wish to edit. You may thus protect against changing the values of any variables except the ones

you select here. The default is to edit all the variables in the dataset.

Once you have selected the dataset and variables which you will be editing, enter NEXT to enter the editor. Once you are in the editor, it behaves as described previously in Section 5.1.2 under the CREATE menu. Please see that section for a description of the editor.

When you exit from the editor, the updates will be performed on the dataset you have edited. Should you quit from the editor (key 9 on the keypad), no modifications will be made to the dataset.

5.2.3 Add Records (ADDREC)

This procedure is used to add additional records to an already existing dataset. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Add Records to a Dataset
ref   parmname     parameter description          value   index
1.    DATASET      Name of the input dataset
2.    RECNUM       Preceding record number          *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 5-14. Add Records to a Dataset Menu

1. DATASET : The name of the dataset to which you will add records.

2. RECNUM : The number of a record in the data file. Enter the record number preceding the one to be entered. (You should enter "0" if the record is to be added at the top.) If you use the default the new records will be added at the end of the dataset.

Enter values for the menu items; enter NEXT to proceed to the next step where the system will display which record number you will be entering and then prompt you for a value for each variable in the dataset. After each record, you are asked if more records will be added. If you answer YES, the loop will repeat; if NO, the update will be performed and a message displayed to inform you of the successful completion of the update.

5.2.4 Delete Records (DELREC)

The Delete Records procedure allows you to delete a data record from an existing dataset in your user dataset library. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Delete Records from a Dataset
ref   parmname     parameter description          value   index
1.    DATASET      Name of the dataset
2.    VARIABLE     Names of the verification variable   *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 5-15. Delete Records from a Dataset Menu

1. DATASET : The name of the dataset from which you will be deleting records.

2. VARIABLE : An optional verification variable. If you specify a variable name here, that variable's value will be displayed before the value is deleted. The default assumes no verification variable.

When you have filled in the menu, enter NEXT to proceed to where you will see:

Enter a record number for record deletion
GEMS >

If you have entered a verification variable, when you enter the record number, GEMS tells you the current data value and asks you to verify that it is deleting the proper record.

Variable AREA in record # 7 = 2.5
Enter YES or NO to confirm the record deletion
Default = YES
GEMS >

Enter YES to delete (or *carriage return*). You will then be asked for the next record number you wish to delete. If you have specified all of the records you would like to delete, enter a carriage return and you will be asked to **Enter GO to begin processing**. This is your last chance to change your mind. Enter NO to abort the record deletion; to delete, enter GO. You will be returned to the MODIFY menu when the deletion is complete.

5.2.5 Modify Dataset Attributes (ATTRIBUTE)

The ATTRIBUTE procedure allows you to modify various attributes of your datasets such as the name, tag field or variable names. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

MENU: Modify Dataset Attributes

ref	parmname	parameter description	value	index
1.	CATEGORY	Data Category	DATASET	
2.	OLDNAME	Current dataset name		
3.	NEWNAME	New dataset name	*	
4.	NEWTAG	Current dataset tag field	*	
5.	OLDVARS	Current dataset variable names	*	(1)
6.	NEWVARS	New dataset variable names	*	(1)

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-16. Modify Dataset Attributes Menu

1. CATEGORY : The GEMS File Management category you will be operating on. These are DATASET, GRAPH, TABLE, MAP or REPORT. See the start of this chapter for an explanation of each data category.

2. OLDNAME : The name of the file for which you will be modifying one or more of the attributes.

3. NEWNAME : The new name which you would like to have the file named. You need only specify this if it is different from the OLDNAME.

4. NEWTAG : The parameter used to change the tag field associated with the file. The tag field is a field of up to 60 characters which can be used to describe the contents of the file for future reference.

5. OLDVARS :
6. NEWVARS : The information for the modifications you would like to make to your dataset, map, graph, or report in the format shown. (Variable names are applicable only to datasets.)

When you have filled the menu, enter NEXT to perform the modifications. The system will display a message informing you of the successful attribute modification and then return you to the modify menu.

5.3 Delete Datasets/Graphs/Tables/ Maps/Reports (DELETE)

The DELETE procedure allows you to clean up your GEMS user dataset library by deleting any old datasets (or graphs...) which you no longer need access to. Remember that on the VAX Cluster, any files which have not been accessed within 45 days will be written to tape and deleted. This includes any GEMS datasets or graphs, etc. which you are maintaining in

your user dataset library. You should use this DELETE procedure to keep your GEMS disk area cleaned up of datasets which you no longer need.

After making a selection from this menu, you will be asked to provide the names of the items to be deleted. You will then be asked to **Enter GO to begin processing**.

This is your last chance to change your mind. Once you enter GO the delete will be performed and you will be returned to the main DELETE menu. The parameter entry menu for this option is shown below.

```
MENU: Delete Datasets/Graphs/Tables/Maps/Reports

1. Delete Datasets                    (DELDATASET)
2. Delete Graphs                      (DELGRAPH)
3. Delete Tables                      (DELTABLE)
4. Delete Maps                        (DELMAP)
5. Delete Reports                     (DELREPORT)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 5-17. Delete Datasets/Graphs/Tables/Maps/Reports Menu

5.4 Output Datasets/Graphs/Tables/ Maps/Reports (OUTPUT)

The OUTPUT procedure allows you to display the information which is contained in a GEMS dataset, graph, table, map, or report. The Review a Dataset and Review a Graph functions will be discussed below. As all of the other functions are equivalent to the Review a Graph function in terms of their user interaction, they will not be discussed specifically.

```
MENU: Output Datasets/Graphs/Tables/Maps/Reports

1. Review a Dataset                    (DATASET)
2. Review a Graph                      (GRAPH)
3. Review a Table                      (TABLE)
4. Review a Map                        (MAP)
5. Review a Report                     (REPORT)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 5-18. Output Datasets/Graphs/Tables/Maps/Reports Menu

5.4.1 Review a Dataset (DATASET)

This procedure allows you to display the values associated with the variables in a GEMS dataset. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Review a Dataset
ref   parmname     parameter description          value   index
1.    DATASET      Name of the dataset
2.    VARIABLS     Names of the output variable    *       (1)

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 5-19. Review a Dataset Menu

1. DATASET : The name of the dataset which is to be reviewed. This can be either a system-installed or a user dataset.

2. VARIABLS : The names of the variables which are to be displayed. Up to 20 variables may be

entered. The default will display all of the variables in the dataset.

To proceed to the next step where the selected variables will be displayed, enter **NEXT**. You cannot display more than 120 characters across the screen. Should you attempt to display more than this, the excess will be displayed on successive screens.

The display will look like the following example:

```
Data List of Dataset: CLIMATE          Number of records = 38700

REC # | MONTH
1     |   1
2     |   2
3     |   3

Press RETURN to page forward, enter Pnnn to position the starting
record of the next page, enter BACK to reselect variables, or enter END
to stop
GEMS >
```

FIGURE 5-20. Review Variables Menu

As you can see, you may now move around within the dataset to display the sections of interest to you. Once you have finished reviewing the dataset, enter **END** and you will be returned to the main OUTPUT menu.

There are several other commands available as you scan a dataset:

PASS : Moves along the output record to display variables further along. Used where the record being displayed has more variables than can be displayed on one screen.

START : When PASS has been used, START allows you to reposition the display at the start of the record.

GROUP N : Displays the specified group. If no group is specified, then the next group is displayed.

FIND : Allows you to search for specific variable values. (FIND VAR_1 = VALUE_1; VAR_2 = VALUE_2...) Up to five variable/value pairs may be specified with the command. The FIND command will search for records which contain the specified variable/values

5.4.2 Review a Graph (GRAPH)

This procedure allows you to review a graph which was created previously. You must be on the device type for which the graph was originally created in order to review the graph.

When you select this option you will see the prompt:

Enter the name of the graph to be reviewed
GEMS >

After giving the graph name, the graph will be displayed at your terminal. After the graph is displayed, you will be returned to the main OUTPUT menu.

Remember, the other review functions (TABLE, MAP, and REPORT) function the same as GRAPH. You simply name the item and it will be displayed for your review.

5.5 Export Datasets/Graphs/Tables/ Maps/Reports (EXPORT)

The EXPORT procedure allows you to output GEMS datasets, graphs, tables, maps or reports from the GEMS file management system to files in your VAX directories to be manipulated outside of GEMS.

This procedure is particularly useful in exporting files to be downloaded to a PC for local processing.

```
MENU: Export a Dataset from the GEMS System

1. Convert a GEMS Dataset to VAX Files          (G2VAX)
2. Convert a GEMS Dataset to SAS Files          (G2SAS)
3. Convert a GEMS Dataset to a .DBF File        (G2DBF)
4. Convert a GEMS Dataset to a .DIF File        (G2DIF)
5. Convert a GEMS Graph to a Graphics File      (G2GRAPH)
6. Convert a GEMS Table to a Table File         (G2TABLE)
7. Convert a GEMS Map to a Map File             (G2MP)
8. Convert a GEMS Report to a Report File       (G2REPORT)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 5-21. Export a Dataset from the GEMS System Menu

5.5.1 Convert a GEMS Dataset to a VAX Format (G2VAX)

This option allows you to convert a GEMS dataset to a flat ASCII VAX file. This can be very useful if you need to print or format the data for a report or download the data to a PC for local processing.

Up to 200 variables from the specified dataset may be specified for conversion, or you can convert the entire dataset. The default "*" value will convert the entire dataset. A header record containing names of all the variables is added to the output file for the VAX. Actual data begins on the second record.

Additionally, a description file named {FILENAME}.DSC will also be created listing the variable name, description and type of each variable written to the exported file. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Convert a GEMS Dataset to VAX Files
ref   parmname     parameter description          value   index
1.    DATASET      Name of the dataset
2.    OUTPUT       Name of the output VAX file
3.    VARIABLE     Names of the output variables   *       (1)

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 5-22. Convert a GEMS Dataset to VAX Files Menu

1. DATASET : The name of the dataset which you are converting to a VAX file.

2. OUTPUT : The name of the VAX file to be created. This can be a string of up to 60 characters. You may include an optional device and directory to write the file to. If you do not specify one, the file will be written to your default directory.

   If you are writing out a multiple group GEMS dataset, you will have several files created. One VAX file will be created for each group in the GEMS dataset. The filenames will have a single digit postfixed to them to indicate which group's data it contains.

3. VARIABLE : You may optionally specify a list of variables to be converted. You may specify up to 200 variables to be written out. If you do not specify anything here, the default is to convert all variables in the dataset.

When you have completed the menu, enter **NEXT** and the procedure will be executed. When it finishes it will display a message: **ASCII file XXX.DAT successfully created, Format**

description file XXX.DSC created and you will be returned to the main EXPORT menu..

5.5.2 Convert a Graph to a Graphics File (G2GRAPH)

This procedure allows you to output a graph from your GEMS user dataset library to a VAX file. This can be useful in allowing you to send the graph to another user or to download the graph to a PC capable of displaying it. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Convert a GEMS Graph to a Graphics File
ref   parmname      parameter description          value   index
1.    GRAPH         Name of the GEMS graph dataset
2.    OUTFILE       Names of the output graphics file

Enter one or more combinations of: reference or parameter name and
valu(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS>
```

FIGURE 5-23. Convert a GEMS Graph to a Graphics File Menu

1. GRAPH : This is the name of the GEMS graph which will be written out to a VAX file.

2. OUTFILE : The name of the VAX file which the graph will be written out to.

Enter the **NEXT** command to execute the procedure when you have completed the menu. When the procedure has been completed, the system will display a message to let you know. It will then return you to the OUTPUT menu.

5.6 Import Datasets/Graphs/Tables/ Maps/Reports (IMPORT)

The IMPORT procedure allows you to take a file in your VAX directory and install it in your user dataset library as the appropriate type of file (dataset, graph, etc.).

```
MENU: Import Datasets/Graphs/Tables/Maps/Reports

1. Convert a VAX File to a GEMS Dataset          (VAX2G)
2. Convert a SAS File to a GEMS Dataset          (SAS2G)
3. Convert a Graphics File to a GEMS Graph       (GRAPH2G)
4. Convert a Table File to a GEMS Table          (TABLE2G)
5. Convert a Map File to a GEMS Map              (MAP2G)
6. Convert a Report File to a GEMS Report        (REPORT2G)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 5-24. Import Datasets/Graphs/Tables/Maps/Reports Menu

Select a procedure from the six listed by entering either the number or the command. These procedures allow you to import files created outside of GEMS into the GEMS catalog as new datasets. These newly created datasets may be used in any of the other operations within GEMS. You must remember that ASCII files which are to be cataloged must have a fixed record format for variable data fields.

We will explore three of the above options that represent the entire list in functionality.

- Convert a VAX File to GEMS Dataset
- Convert a SAS File to GEMS Dataset
- Convert a Graphics File to GEMS Dataset

The Convert a Map File is the same as the Convert a Graphics File, and the remaining procedures are the same as the Convert a SAS File, so they will not be discussed individually.

5.6.1 Convert a VAX File to a GEMS Dataset (VAX2G)

This procedure converts a VAX file to a GEMS dataset in your user dataset library. The VAX file which this procedure accesses must be a fixed format file. In other words, the data fields must have fixed positions on each record in the file. You may create this file by any means available including the VMS editor. The GEMS dataset which this procedure creates may be accessed by all other GEMS functions. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

MENU: Convert a VAX File to a GEMS Dataset

ref	parmname	parameter description	value	index
1.	INSET	Name of the input VAX file		
2.	NHEADERS	Number of header records		(1)
3.	OUTSET	Name of the output GEMS dataset		
4.	TAG	Tag field of the output dataset	*	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-25. Convert a VAX File to a GEMS Dataset Menu

1. INSET : The name of the input VAX file. This is the file from which the new GEMS dataset will be created.

2. NHEADERS : The number of records in the file which should be skipped before the conversion process is started. This allows you to place some textual information or field markers in the file while creating it to help yourself out, but have GEMS ignore that

information while creating the new dataset.

3. OUTSET : The name of the new dataset which will be created from the input VAX file.

4. TAG : A 60 character description field.

Complete the menu and enter NEXT to proceed to the next step where you will see:

Enter the name of the 1st data field in the VAX file

The name that you enter will be the name of the variable in the GEMS dataset. You will also be asked to identify the type of field (I)nteger, (R)eal, (C)haracter, and the length of the field. This process will be repeated until you have specified name, type, and length for all fields in the new dataset. Once all fields have been defined, the VAX file conversion will begin. When the process is complete, the system will display the following message and return you to the IMPORT menu.:

GEMS dataset {OUTSET} successfully created

5.6.2 Convert a SAS File to a GEMS Dataset (SAS2G)

This procedure allows you to convert a SAS file to a GEMS dataset. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Convert a SAS File to a GEMS Dataset
ref   parmname     parameter description          value   index

1.    INSET        Name of the Input SAS file
2.    OUTSET       Name of the output GEMS dataset
3.    TAG          Tage field of the output dataset     *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 5-26. Convert a SAS File to a GEMS Dataset Menu

1. INSET : The filename of the SAS file to be converted.

2. OUTSET : The name for the new GEMS dataset which will be created by this procedure.

3. TAG : A description field of up to 60 characters for the new dataset.

The variable names and field sizes will be retrieved from the SAS file which is being converted. Once your information is entered in the indicated format, use the NEXT command to complete the conversion of the file. When the conversion has been completed, the system will display a message:

GEMS dataset {OUTSET} successfully created.

5.6.3 Convert a Graphics File to a GEMS Dataset (GRAPH2G)

This procedure allows you to enter a graphics file into your GEMS dataset library. The parameter entry menu for this option follows, and a discussion of each parameter follows.

MENU: Convert a Graphics File to a GEMS Graph

ref	parmname	parameter description	value	index
1.	INFILE	Name of the graphics file		
2.	FORMAT	File format: SAS or VAX		
3.	TERM	Output terminal type: TEK or ASCII		
4.	GRAPH	Name of the output graph dataset		
5.	TAG	Tag field of the output graph dataset	*	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 5-27. Convert a Graphics File to a GEMS Graph Menu

1. INFILE : The name of an existing VAX graphics file to be imported into the GEMS file management system.

2. FORMAT : Either of two GEMS graphics file formats: SAS and VAX. A SAS graphics file is created by a SAS graphical procedure, while a VAX graphics file contains escape sequences which may be created by PLOT10 or FORTRAN I/O routines.

3. TERM : The terminal device required for the graphics display may be either a Tektronix device (TEK) or an ASCII terminal.

4. GRAPH : The name for the graph in your graph library.

5. TAG : This is a 60 character description of what the graph is displaying.

After entering your information, only the NEXT command is required to complete the process. Once finished, the system will display a message to let you know that the graph was successfully created.

5.7 Change Status of Datasets/Graphs/ Tables/Maps/Reports (STATUS)

These procedures allow you to make a temporary dataset (or graph, etc.) permanent. Remember that temporary datasets are deleted automatically when you exit from GEMS while permanent datasets will be retained for later use.

Datasets not accessed for 45 days will be automatically archived by TAPESYS on the VAX Cluster. You should delete any unnecessary datasets before this happens.

```
MENU: Change Status of Datasets/Graphs/Tables/Maps/Reports

1. Save Temporary Datasets                    (SAVEDATASET)
2. Save Temporary Graphs                      (SAVEGRAPH)
3. Save Temporary Tables                      (SAVETABLE)
4. Save Temporary Maps                        (SAVEMAP)
5. Save Temporary Reports                     (SAVEREPORT)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 5-28. Change Status Menu

All of the SAVE functions work exactly the same so only the SAVEDATASET function will be described below.

5.7.1 Save Temporary Dataset (SAVEDATASET)

This option allows you to save selected datasets in your user dataset library. When it is selected, you will be asked:

Enter the name of the 1st temporary dataset to be saved
GEMS >

After you enter the name, you will be asked to name additional datasets to be saved. Enter all of the datasets that you would like to save, followed by a carriage return. The dataset entries will be changed to reflect their new status. When the procedure ends, you will be returned to the main menu for this section.

The other four SAVE functions are identical to the Save Dataset function. You need only substitute the name of the item to be saved.

6 - Statistics

The STATISTICS operation gives you the ability to run a variety of statistical analyses on the numeric variables contained in GEMS datasets. This procedure uses library routines from the International Mathematical and Statistical Library (IMSL) and the Statistical Analysis System (SAS) to perform its computations.

The statistical procedures available in GEMS are defined by the type of procedure and are displayed in the STATISTICS main menu.

```
MENU: Statistics

1. Contingency Tables                    (CT)
2. Descriptive Statistics                (DS)
3. Simple Regression                     (SR)
4. Multiple Regression                   (MR)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 6-1. Statistics Menu

6.1 Contingency Tables (CT)

This procedure enables you to tally observations into a two-way contingency table on two variables at a time and perform a contingency table analysis. Each variable is divided

into three equal ranges between the maximum and the minimum of that variable. The chi-square analysis is performed on the table. The chi-squared value, degrees of freedom, the mean and standard deviation of the chi-squared are provided. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Contingency Tables
ref   parmname     parameter description          value   index
1.    DATASET      Name of the Input Dataset
2.    VAR1         Name of the First Variable
3.    VAR2         Name of the Second Variable
4.    TABLE        Name of the Output Table
5.    TAG          Tag Field of the Output Table    *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 6-2. Contingency Tables Menu

1. DATASET : Name of the dataset containing the variables to be used in the contingency table.

2. VAR1 : First variable to be used in the contingency table.

3. VAR2 : Second variable to be used in the contingency table.

4. TABLE : Name of the output table being created by this procedure.

5. TAG : Description field for the output table being created by this procedure.

Fill in all menu entries and enter NEXT to execute the procedure. The output will be stored as a table in your GEMS dataset library.

6.2 Descriptive Statistics (DS)

This procedure provides access to statistical routines which perform statistical analyses on the variables contained in GEMS datasets. These procedures compute statistics on the distribution of the numerical variables of the dataset.

```
MENU: Descriptive Statistics

1. Simple Statistics                    (MEAN)
2. Univariate Statistics                (UNIVARIATE)
3. Percentile Statistics                (PERCENTILE)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 6-3. Descriptive Statistics Menu

6.2.1 Simple Statistics (MEAN)

This procedure computes distribution statistics such as the mean, standard deviation, variance, minimum value, maximum value, sum, coefficient of variation, and N (the number of non-missing values). The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

```
MENU: Simple Statistics
ref   parmname     parameter description                  value  index
1.    DATASET      Name of the Input Dataset
2.    VARIABLE     Name of the Input Numeric Variables     (1)
3.    TABLE        Name of the Output Table
4.    TAG          Tag Field of the Output Table

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 6-4. Simple Statistics Menu

You may specify more than one variable for this procedure.

1. DATASET: Name of the dataset containing the variable from which the statistics will be computed.

2. VARIABLE: Variable(s) for which the statistics will be generated.

3. TABLE: Name of the output table being created.

4. TAG: Description field for the output table being created by this procedure.

Complete the menu and enter NEXT to execute the procedure and display the results. This is followed by:

Table {table} successfully completed.

GEMS will substitute that name of output table that you entered for {table}. When the procedure is completed, you will be returned to the DESCRIPTIVE STATISTICS menu (Figure 6-3).

6.2.2 Univariate Statistics (UNIVARIATE)

This procedure is a more comprehensive analysis on the variables, providing order statistics in addition to those provided by the mean procedure. These include quartiles, highest and lowest five, kurtosis, and skewness measures. The parameter entry menu for this option is shown below.

MENU: Univariate Statistics

ref	parmname	parameter description	value	index
1.	DATASET	Name of the Input Dataset		
2.	VARIABLE	Name of the Input Numeric Variables		(1)
3.	TABLE	Name of the Output Table		
4.	TAG	Tag Field of the Output Table		

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 6-5. Univariate Statistics Menu

1. DATASET : Name of the dataset containing the variable to use.

2. VARIABLE : Variable for which the statistics will be generated.

3. TABLE : Name of the output table being created.

4. TAG : Description field for the output table being created by this procedure.

Complete the menu and enter NEXT to execute the procedure and display the results. This is followed by:

Table {table} successfully completed.

GEMS will substitute the name of the output table that you entered for {table}. When the procedure is completed, you will be returned to the DESCRIPTIVE STATISTICS menu.

6.2.3 Percentile Statistics (PERCENTILE)

This routine produces equidistant percentiles and maximum and minimum values on a single numeric variable. You can choose the number of percentiles that are desired, e.g., if you choose four percentiles, the routine will generate quartiles. The parameter entry menu for this option is shown below, and a discussion of each parameter follows.

MENU: Percentile Statistics

ref	parmname	parameter description	value	index
1.	DATASET	Name of the Input Dataset		
2.	VARIABLE	Name of the Input Numeric Variables		(1)
3.	LEVELS	Number of the Percentile Levels		
4.	TABLE	Name of the Output Table		
5.	TAG	Tag Field of the Output Table		

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 6-6. Percentile Statistics Menu

1. DATASET : Dataset containing the variable from which to generate the percentile statistics.

2. VARIABLE : Variable for which the statistics will be generated.

3. LEVELS : Number of percentile divisions for which the statistics are generated.

4. TABLE : Name of the output table being created by this procedure.

5. TAG : Description field for the table being created by this procedure.

Complete the menu and enter NEXT to execute the procedure and display the results. The results are followed by:

Table {table} successfully completed.

GEMS will substitute the name of the output table that you entered for {table}. When the procedure is completed, you will be returned to the DESCRIPTIVE STATISTICS menu.

6.3 Simple Regression (SR)

The Simple Regression procedure enables you to perform a linear regression analysis of two variables. The statistics the procedure produces include:

1. Mean of each variable.
2. Standard deviation.
3. Degrees of freedom for the regression, residual and corrected totals sources of variation.
4. Sums of squares corresponding to the degrees of freedom in 3.
5. Mean squares for the regression and residual sources of variation.
6. F-value.
7. Percentage of variation explained.
8. Regression coefficient.
9. Standard error of the regression coefficient.
10. Estimate of the intercept.
11. Standard error of the intercept.

MENU: Simple Regression

ref	parmname	parameter description	value	index
1.	DATASET	Name of the Input Dataset		
2.	INDEPVAR	Name of the Independent Variable		
3.	DEPVAR	Name of the Dependent Variable		
4.	TABLE	Name of the Output Table		
5.	TAG	Tag Field of the Output Table	*	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 6-7. Simple Regression Menu

1. DATASET: Name of the GEMS dataset containing the variables to be analyzed.

2. INDEPVAR : Variable by which the specified dependent variable will be analyzed. This variable must exist in the input dataset.

3. DEPVAR : Variable to be analyzed as a function of the independent variable. This variable must exist in the input dataset.

4. TABLE : Name of the output table being created by this procedure.

5. TAG : Description field of up to 60 characters which you may enter to help identify the table in the future.

Fill in all menu entries and enter NEXT to execute the procedure. The output will be stored as a table in your GEMS dataset library.

6.4 Multiple Regression (MR)

The Multiple Regression procedure allows you to perform a linear regression analysis of up to four, independent variables on one dependent variable. This procedure produces the following regressions and analyses of variance statistics.

1. Mean of each variable.
2. Degrees of freedom for the regression, residual and corrected total.
3. Sums of squares corresponding to the degrees of freedom in 2.
4. Mean squares for the regression and residual sources of variation.
5. F-value.
6. Percentage of variation explained.

7. Standard error of the parameters estimate.
8. Partial F-value.

```
MENU: Multiple Regression
ref   parmname     parameter description              value  index
1.    DATASET      Name of the Input Dataset
2.    INDEPVAR     Name of the Independent Variable
3.    DEPVAR       Name of the Dependent Variable
4.    TABLE        Name of the Output Table
5.    TAG          Tag Field of the Output Table        *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

FIGURE 6-8. Multiple Regression Menu

1. DATASET : Name of the GEMS dataset containing the variables to be analyzed.

2. INDEPVAR : Variable by which the specified dependent variable will be analyzed. This variable must exist in the input dataset.

3. DEPVAR : Variable to be analyzed as a function of the independent variable. This variable must exist in the input dataset.

4. TABLE : Name of the output table being created by this procedure in your user dataset library.

5. TAG : Description field of up to 60 characters which you may enter to help identify the table in the future.

Complete all menu entries and enter *NEXT* to execute the procedure. The output will be stored as a table in your GEMS dataset library.

Chapter 7 - Graphics

The GEMS Graphics operation provides a variety of graphic procedures for use in displaying results of GEMS modeling routines and datasets.

Graphics functions available in GEMS include:

- Line plotting with linear and nonlinear regression
- bar charts
- histograms
- scattergrams
- isopleths
- rose diagrams

GEMS depends on the Statistical Analysis System (SAS) for much of the graphic display and mapping capabilities. The SAS-GEMS tandem provides powerful graphics support for Tektronix graphics terminals, VT100 terminals with TEK 4010 emulation, and for PCs with Tektronix emulation software.

As of the writing of this chapter, only node VAXTM1 of the EPA VAX Cluster can use SAS. GEMS features which use SAS can, therefore, only be executed on this node which is VAXA on the NCC port selector. If you do attempt to use one of the graphics operations on a node other than VAXTM1, you will be warned after the first screen and returned to the GEMS Graphics menu.

Assuming you have already logged-on to GEMS from the main menu, the GR command takes you to the GEMS Graphics menu.

```
MENU: Graphics

1. Barchart                    (BAR)
2. Histogram                   (HIST)
3. Scattergram                 (SCAT)
4. Isopleth                    (ISO)
5. Rose                        (ROSE)

Enter an option number or a procedure name (in parenthesis) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 7-1. Graphics Menu

Each of the above graphics procedures will be discussed in detail in the following sections.

7.1 Barchart (BAR)

The Barchart procedure produces numeric-variable and character- variable horizontal and vertical bar charts. The bar charts show relationships between various midpoint levels and are useful for exploratory data analysis. The vertical bar charts are similar to horizontal bar charts, except that statistics are not printed for each bar.

Six types of statistics may be collected and displayed for each midpoint:

- Frequency counts - the number of observations contained in each midpoint level.
- Cumulative frequency counts - the number of observations at a specific midpoint level, including observations from all preceding midpoint levels.

- Percentages - the number obtained when you divide each interval level frequency by the total number of scores and multiply the result by 100.

- Cumulative percentages - the number obtained when you divide each interval level frequency by the total number of scores, multiply the result by 100, and include the percentages from all preceding midpoint values.

- Sums - the midpoint value derived from another variable in the data.

- Means - the number obtained when you sum a list of score values and divide the number of scores used in the calculation.

When the Graphics menu appears, you may enter either 1 or BAR to begin the Barchart procedure. The navigational menu shown below appears.

```
MENU: Barchart (Output Selection)

1. Graphic Output Barchart                    (BARGRAPH)
2. Character Output Barchart                  (BARCHAR)


Enter an option number or a procedure name (in parenthesis) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 7-2. Barchart Options Menu

The character output option uses alphabetic characters to draw the bars and patterns and may be used with any terminal. The graphic output option uses graphics characters, and is appropriate only when you are using a Tektronix graphic terminal or PCs with Tektronix emulation software.

The next menu you see will allow you to define the specifics of the chart and establish the output parameters. This menu will have one of two different titles, depending on which option you

selected above, but the parameters to be specified will be identical. The parameter entry menu that will appear if you select the character output option is shown below, followed by a discussion of each parameter.

MENU: Barchart (character output)

ref parmname	parameter description	value	index
1. DATASET	Dataset name		
2. GRAPH	Output graph name		
3. TAG	Description of Tag Field	*	
4. PAGESIZE	Number of lines of output page	*	
5. TYPE	Barchart function type (SUM, MEAN)	SUM	
6. MSVAR	Sum or Mean variable name		
7. BYVARBY	Variable name		

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,..] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 7-3. Barchart Options (Character Output)

1. DATASET: The name of the dataset which contains the variables you will use as input.

2. GRAPH: The name of the barchart that you are creating. It will be stored in your GEMS User Library.

3. TAG: A descriptive field of up to 60 characters to help you identify the graph in the future.

4. PAGESIZE: The number of lines available on your terminal for the graph. This parameter will only appear on the menu if you have selected the character output barchart. Defaults are available for this parameter and are set by the terminal type which you specified when you entered GEMS:

VT100 and compatibles and all TEKxxxx except TEK4014:	24 lines
TEK4O14:	32 lines
LA120:	66 lines
132 ASCII terminal:	24 lines

You may accept the default which the procedure will select for you or set your own line count.

5. TYPE: The method of computing the value of each bar. There are two types available, you must select one.

SUM - sums the values of the MSVAR for each different value or range of the BYVAR. (See definitions below for MSVAR and BYVAR.)

MEAN - computes the mean of the values of the MSVAR for each different value or range of the BYVAR. The procedure uses a default of the SUM function.

6. MSVAR: The variable in the selected dataset to be used to compute the SUMS or MEANS to determine the length of each bar. This variable must be a numeric variable.

7. BYVAR: The variable which defines the ranges in which each of the MSVAR variables will be placed when computed.

If you need help identifying datasets or variables, enter either the LIST DATASET or LIST VARIABLES commands. Once this menu has been filled in, enter NEXT to move to the next step. If you are on a node other than VAXTM1, GEMS will display a warning message and end the procedure. This is because this graphics procedure uses SAS which is available only on VAXTM1.

You will be asked:

Do you want to specify ranges for the BYVAR variable (Y,N)
GEMS >

Each bar in the barchart will represent a range of values for the BYVAR variable. Enter Y (YES) to specify ranges yourself or N (NO) to allow the procedure to break the BYVAR variable down into ranges. Specifying ranges allows you to select the range of values and a label for this range.

If you answered YES to this prompt, you will be asked to:

Enter the upper limit of the 1st range
GEMS >

You will then be asked the following:

Enter a label for the 1st range
GEMS >

This section will repeat until you enter a carriage return for one of the upper limits. You will then be asked:

Enter the label for the end range
GEMS >

This will provide a label for a bar representing all values exceeding the maximum one you specified in your previous range limits.

If you answered NO to the prompt as to whether you wanted to specify ranges, you will be asked:

Does this variable have discrete values? (Y,N)
GEMS >

The definition of a discrete variable is that it is a mathematical variable which assumes only whole number value. For example, a variable with the values 1, 2, 3, and 4 is a discrete

variable while a variable with the values 1.5, 2.5 and 3 is not a discrete variable. Enter **Y** if the variable has discrete values or **N** otherwise.

Next, you will be asked to identify a final set of barchart options:

```
MENU: Options (Barchart)

ref parmname   parameter description              value  index
1. BARTYPE     Horizontal or Vertical bars (H,V)   V
2. SUBGROUP    Do you want to specify subgroup?    N
3. ORDER       Do you want the bars to be ordered? N
4. AXES        Do you want to specify axes/labels  N
5. TITLE       Do you want to specify titles?      N
6. FTNOTE      Do you want to specify footnotes?   N

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

Figure 7-4. Barchart Options Menu

1. BARTYPE: The type of bar which the procedure will draw. Enter a V for vertical bars or a H for horizontal bars.

2. SUBGROUP: The parameter used to further split the ranges you have defined into subgroups within those ranges. (This parameter will only appear if you have specified ranges for the BYVAR yourself.)

3. ORDER: The parameter used to order the bars in either ascending or descending order. The default is no order to the bars. A "Y" response will activate the option.

For AXES, TITLE, and FTNOTE, you will enter text to specify the appropriate titles and labels. When you have finished specifying the items you have selected, enter **NEXT** and the bar chart will be drawn. An example of a bar chart is provided in Figure 7-5.

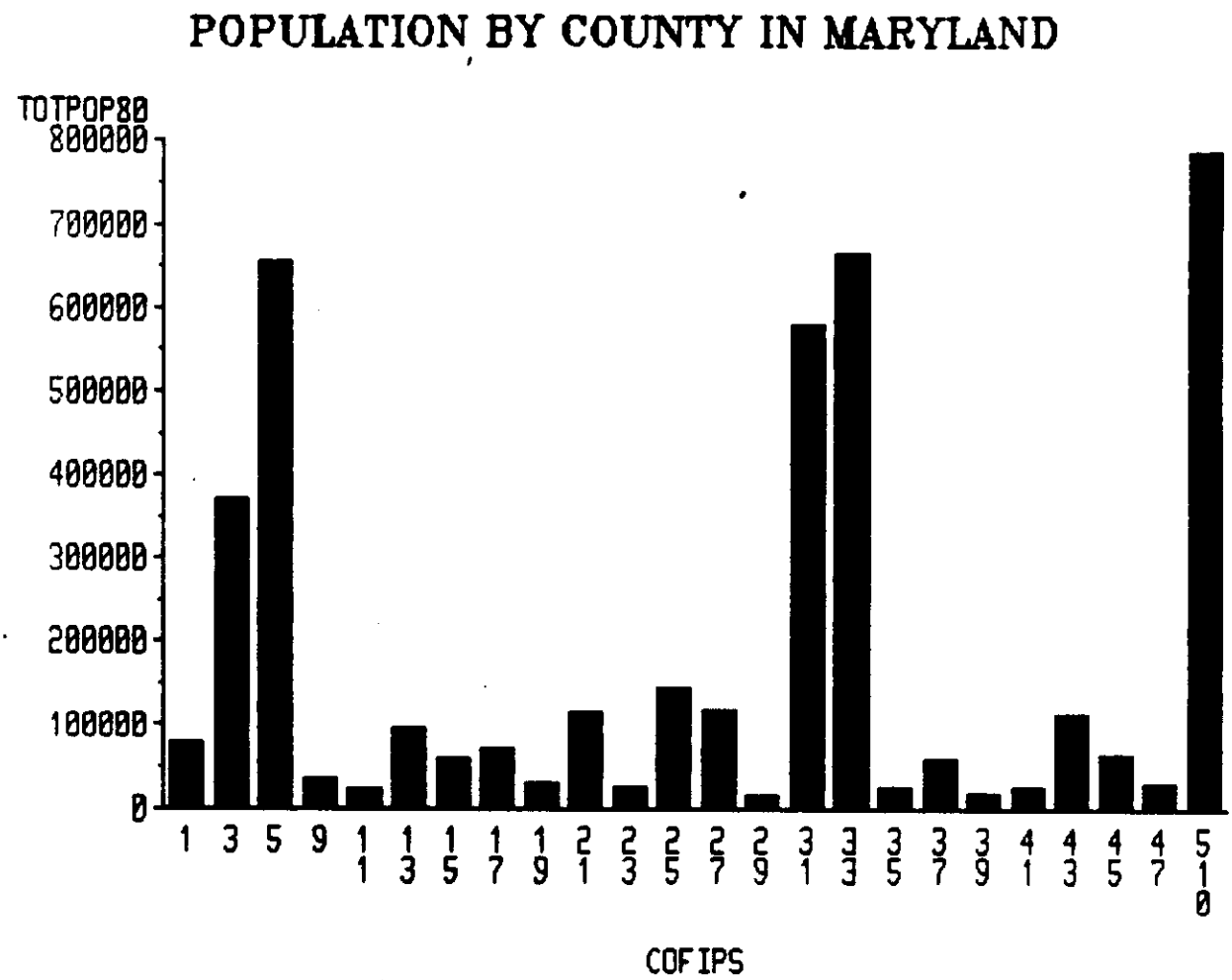

FIGURE 7-5. Example of a Barchart

7.2 Histogram (HIST)

The Histogram procedure produces a graphic representation of grouped data by frequency or percentage. The horizontal axis is a scale of the values of the measured variable, explicitly marked. The vertical axis is the frequency or percentage scale. Above the horizontal axis scale, a rectangular bar is placed with its height corresponding to its measured frequency or percentage. Histograms display the relative frequencies or percentages of occurrences of various values by comparing relative areas.

Enter 2 or the HIST command from the main graphics menu to access the Histogram procedure. You have two output options. Character output uses alphabetic characters to draw the histogram. The graphic output option uses graphic characters, and is appropriate only if you are using a Tektronix graphic terminal or PCs with Tektronix emulation software.

The next menu you will see allows you to define the specifics of the histogram and establish the output parameters. This menu will have one of two different titles, depending on the output option that you select, character or graphic, but the parameters to be specified will be identical. The parameter entry menu for the character output option is shown below.

MENU: Histogram (character output)

ref	parmname	parameter description	value	index
1.	DATASET	Dataset name		
2.	GRAPH	Output graph name		
3.	TAG	Description of Tag Field	*	
4.	PAGESIZE	Number of lines of output	*	
5.	TYPE	Histogram function (FREQ,PCNT)	FREQ	
6.	HISTVAR	Variable to use for histogram		

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

Figure 7-6. Histogram Output Parameter Menu

1. DATASET: The name of the dataset which contains the variables you will use as input.

2. GRAPH: The name that you wish to give to the histogram that is to be produced.

3. TAG: A descriptive field of up to 60 characters to help you identify the graph in the future.

4. PAGESIZE: The number of lines available on your terminal for the graph. This parameter will only appear on the menu if you have selected the character output histogram. Defaults are available for this parameter and are set by the terminal type which you specified when you entered GEMS:

VT100 and compatibles and all TEKxxxx except TEK4014:	24 lines
TEK4O14:	32 lines
LA120:	66 lines
132 ASCII terminal:	24 lines

You can accept the default which the procedure will select for you or set your own line count.

5. TYPE: The method of computing values for the bars. There are two different functions which you can apply to compute the values for the bars. You must select one type.

FREQ - computes the frequency of each distinct value of the variable that is to be used to produce the histogram (called HISTVAR, see definition below).

PCNT - computes the percentage of each distinct value of the HISTVAR.

6. HISTVAR: The variable in the selected dataset which is to be used to produce the histogram. This variable must be a numeric variable.

If you need help identifying datasets or variables, enter either LIST DATASETS or LIST VARIABLES. If you are on a node other than VAXTM1, GEMS will display a warning message and end the procedure. This is because this graphics procedure uses SAS which is available only on VAXTM1.

Once you have entered the specified parameters, you then use the NEXT command to proceed to the next step where you will be asked:

Does the HISTVAR have discrete values? (Y,N)
GEMS >

The definition of a discrete variable is that it is a mathematical variable which assumes only whole number value. For example, a variable with the values 1, 2, 3, and 4 is a discrete variable while a variable with the values 1.5, 2.5 and 3 is not a discrete variable. Enter Y (YES) here if your variable is a discrete variable.

If you enter N (NO), you will be asked to specify the ranges to be used in computing the histogram.

Enter the upper limit of the 1st range
GEMS >

You may enter a carriage return here if you want the system to enter the ranges. If you enter a value, you will be asked:

Enter a label for the 1st range
GEMS >

This section will repeat until you enter a carriage return for one of the upper limits. You will then be asked:

Enter the last label for the ranges
GEMS >

This will provide a label for a bar representing all values exceeding the maximum one you specified in your previous range limits.

Next, you will be asked to identify a final set of histogram options. Note that you may not have some of these parameters on the menu which you see. This menu varies depending on the options you selected prior to this (involving character or graphic bar chart or discrete values of your data). All of the possible entries are detailed here in order to give a full picture of the procedure.

MENU: Options (Histogram)

ref	parmname	parameter description	value index
1.	HISTTYPE	Horizontal or Vertical bars (H,V)	V
2.	BARSPACE	Do you want space between bars?	Y
3.	SUBGROUP	Do you want to specify subgroup?	N
4.	CUM	Do you want a cumulative histogram?	N
5.	AXES	Do you want to specify axes/labels?	N
6.	TICKMARK	Do you want to specify tick marks?	N
7.	PATTERN	Do you want to specify patterns?	N
8.	TITLE	Do you want to specify titles?	N
9.	FTNOTE	Do you want to specify footnotes?	N

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

Figure 7-7. Histogram Options Menu

1. HISTTYPE: The variable used to specify bar direction. Enter an H for horizontal or V for vertical bars.

2. BARSPACE: The variable which specifies whether the bars are to be evenly spaced with intervals in between or grouped together. (This option will only appear if you have selected a graphic bar chart.)

3. SUBGROUP: The variable which allows you to divide the ranges you specified into subgroups to break the display down further. This option only appears if you have discrete values.

4. CUM: The variable which specifies whether you want a histogram where each bar represents only the values falling within its range or whether the histogram should be cumulative of all previous values.

5. AXES: A yes/no switch for axes labels. Set this value to Y to specify labels for the axes of the histogram.

6. TICKMARK: A yes/no switch for tick mark scaling. Enter Y if you want to specify tick marks of the scaled axis. Leave the value at N if you want the tick marks to be defaulted. You can use tick mark specification to exclude values above or below a range that has been specified.

7. PATTERN: A yes/no switch for changing default colors and patterns. Specifying patterns allows you to change the color and pattern of the bars to be drawn. Normally when there are no subgroup specifications, only one color and pattern is used for the bars in the graph. You can change the default pattern and color by specifying Y. When subgroups are specified, every value or range of the HISTVAR variable will be given a different pattern. If this is the case, you may specify as many patterns as you wish (but only as many patterns as needed will be used).

8. TITLE: A yes/no switch for title specification. Change to Y if you would like to specify titles for the histogram.

9. FTNOTE: A yes/no switch for footnotes. Change to Y if you would like to specify footnotes for the histogram.

For each option you select, you will be prompted to enter the appropriate data. After specifying all of the items which you have selected, enter NEXT to draw your graph. An example of a graphic histogram where the data is represented according to frequency is given in Figure 7-8.

7.3 Scattergram (SCAT)

The Scattergram procedure graphs the values of two or more variables from each record in an input dataset and represents the intersection of these values as points on a plot. Using this program, you may also overlay linear or nonlinear regression lines on top of the scattered points. The Scattergram procedure can also superimpose two or more plots, reverse the order of values on the vertical axes so that they decrease rather than increase, and generate a second vertical axis.

From the GEMS Graphics menu, enter either 3 or SCAT to begin the Scattergram procedure. You have two output options. Character output uses alphabetic characters to draw the scattergram. The graphic output option uses graphics characters, and is appropriate only when you are using a Tektronix or Tektronix compatible terminal.

The next menu you see depends on the type of scattergram you selected. If you selected the character output scattergram, you will see the navigational menu shown below:

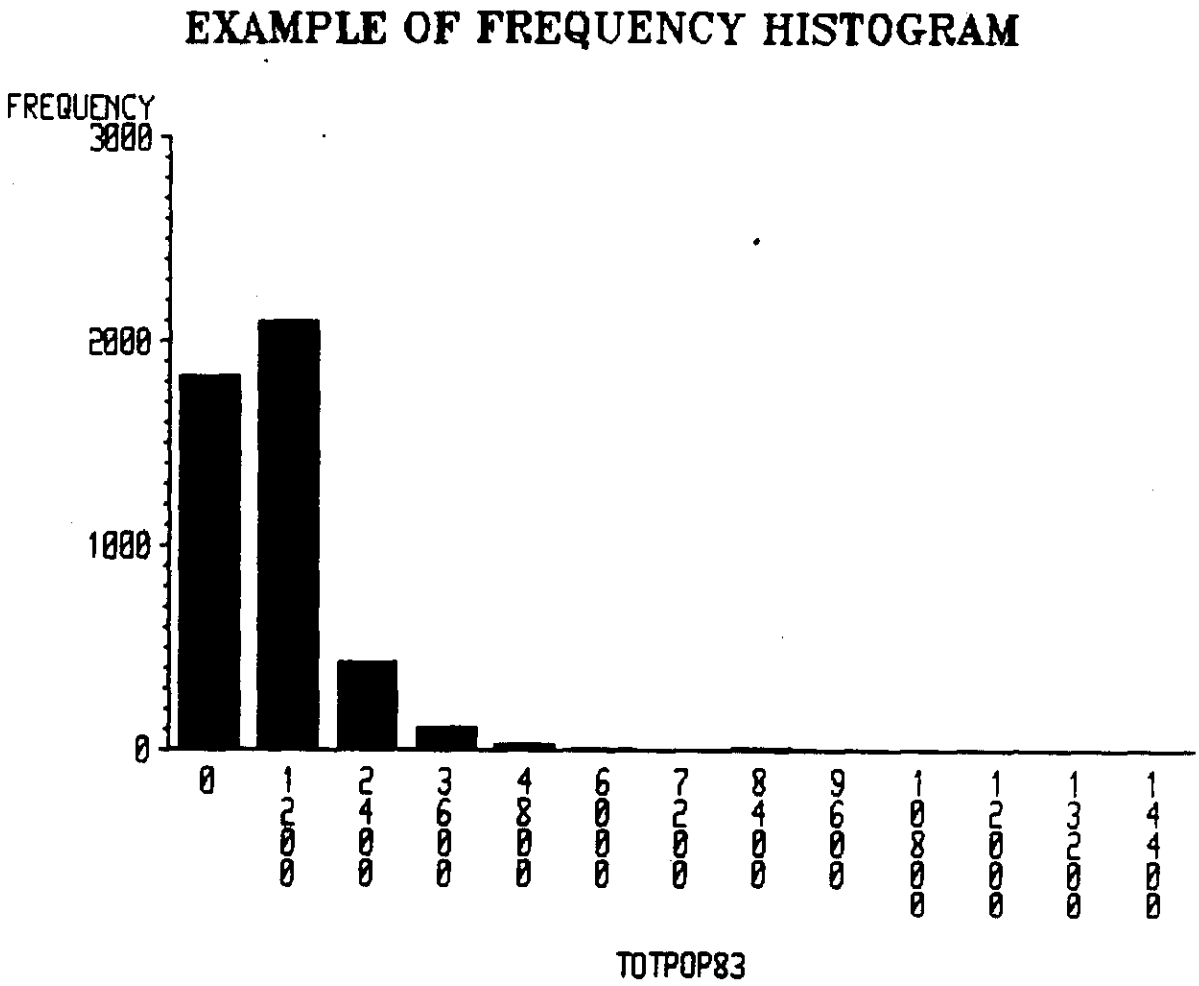

Figure 7-8. Example of Histogram

```
MENU: Character Output Scattergram

1. Plot one variable against another                (SINGLE)
2. Predict and plot against actual values           (LINREG)
3. Overlay of two or more plots (MAX, 4)            (COVERLAY)

Enter an option number or a procedure name (in parenthesis) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR

GEMS >
```

Figure 7-9. Character Output Scattergram

And if you selected the graphic output scattergram, you will see:

```
MENU: Graphics Output Scattergram

1. Plot one variable against another                (SINGLE)
2. Two plots with LEFT and RIGHT vertical axes      (PLOT2)
3. Overlay of two or more plots (MAX, 4)            (OVERLAY)

Enter an option number or a procedure name (in parenthesis) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

Figure 7-10. Graphic Output Scattergram

Depending on which of the plot types you select, you will face a different set of menus. However, the OVERLAY and COVERLAY paths are very similar and will be discussed together under OVERLAY.

7.3.1 Single Plot

If you select the single plot scattergram, the parameter entry menu shown below appears.

```
MENU: Single Plot (One variable against another)

ref parmname   parameter description                value   index
1. DATASET     Name of the output dataset
2. GRAPH       Name of the output graph
3. TAG         Tag Field for the output graph       *
4. XVAR        X-Axis Variable name
5. YVAR        Y-Axis Variable name

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

Figure 7-11. Single Plot Menu

1. DATASET: The name of the dataset which contains the variables you will use as input.

2. GRAPH: The name the scattergram you produce will be stored under in your GEMS User Library.

3. TAG: A descriptive field of up to 60 characters to help you identify the graph in the future.

4. XVAR: The name of the variable for the X axis.

5. YVAR: The name of the variable for the Y axis.

If you are using the character graphic output mode, you will also be asked to specify a PAGESIZE parameter. Use the parameter to specify the number of lines that your terminal can display. If you need help identifying datasets or variables, use the LIST DATASETS or the LIST VARIABLES commands. If you are on a node other than VAXTM1, GEMS will display a warning message and end the procedure. This is because this graphics procedure uses SAS which is available only on VAXTM1.

Once you are satisfied with your designation of the specified parameters, use of the NEXT command takes you to the interpolation methods menu shown below. This menu is only used by the graphics output scattergram and allows you to specify whether to leave the plotted points unconnected or to connect the plotted points with straight lines or a smoothed line.

```
MENU: Select one of the following INTERPOLATION METHODS

1. Leave the points unconnected                    (NONE)
2. Join the points with a straight line            (JOIN)
3. Join the points using CUBIC SPLINE method       (SPLINE)
4. Use a regression analysis equation              (REGRESS)

Enter an option number or a procedure name (in parenthesis) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

Figure 7-12. Interpolation Methods Menu

The JOIN option requests that the points on the plot be connected by a straight line. The SPLINE option specifies that the plot line be interpolated using a spline routine and allows you to smooth points in a plot. In the SPLINE option, the plot line is smoothed using a cubic spline method. If you select the REGRESS option, you will then be asked to specify Linear, Quadratic, or Cubic regression. You will also have the opportunity to force the regression line through the origin, or specify a confidence level of NONE, 90, 95, or 99. You can also place a limit on the MEAN or INDIVIDUAL predicted values.

General options available to you are then presented in the parameter entry menu shown below.

The parameters in this menu have already been defined in section 7.2. If you have previously specified character output rather than graphic output, this general options menu will differ. The SPECIFY parameter will allow you to specify plot symbols, rather than lines and colors. The FTNOTE function is not supported for character output.

```
MENU: General Options

ref parmname    parameter description                  value index
1.  SPECIFY     Plot characters, lines, and colors?    N
2.  AXES        Do you want to specify Axes/Labels?    N
3.  TICKMARK    Do you want to specify tick marks?     N
4.  TITLE       Do you want to specify titles?         N
5.  FTNOTE      Do you want to specify footnotes?      N

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

Figure 7-13. General Options Menu

To select an option, change that options value to Y. For each option you select, you will be prompted to enter the appropriate data in a subsequent menu. For the AXES parameter, you will have to enter labels for X and Y and the reverse axes. After specifying all of the options you have selected, enter *NEXT* t o proceed to draw the graph.

7.3.2 Plot2

Option 2 from the plot selection menu (graphics output option) takes you to the Plot2 Specifications Menu.

```
MENU: Plot2 (Left and Right Vertical Axes)

ref parmname    parameter description                  value index
1.  DATASET     Name of the input dataset
2.  GRAPH       Name of the output graph
3.  TAG         Tag field for the output graph         *
4.  XVAR        X-Axes variable name
5.  YLVAR       LEFT Y-Axes variable name
6.  YRVAR       RIGHT Y-Axes variable name

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
```

Figure 7-14. Plot 2 Specifications Menu

1. DATASET: The name of the dataset which contains the variables you will use as input.

2. GRAPH: The name that you wish to give the scattergram.

3. TAG: A descriptive field of up to 60 characters to help you identify the graph in the future.

4. XVAR: The name of the variable for the X axis.

5. YLVAR: The name of the variable for the left Y axis.

6. YRVAR: The name of the variable for the right Y axis.

If you are operating in the character mode, a seventh option, "PAGESIZE" is available. This option allows you to specify the length, in lines, for your output page. However, default values (which you may assume) are automatically assigned for this variable according to the terminal type you specified.

Once you have finished, the NEXT command tells the system to validate your datasets and variables and proceed to the next step. If you have specified graphic output scattergram, then you must specify the interpolation method, in the menu shown below. If you selected character output scattergram, the only option is NONE and you will skip this menu and proceed to the next step.

```
MENU: Select one of the following INTERPOLATION METHODS:

1. Leave the points unconnected                (NONE)
2. Join the points with a straight line        (JOIN)
3. Join points using CUBIC SPLINE method       (SPLINE)
4. Use a regression analysis equation          (REGRESS)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

Figure 7-15. Interpolation Methods Selection Menu

See the discussion in section 7.3.1 about interpolation methods. When you have selected an interpolation method, the following parameter entry menu appears. If you have previously specified character output rather than graphic output, this general options menu will differ. The SPECIFY parameter will allow you to specify plot symbols rather than lines and colors. The FTNOTE function is not supported for character output.

```
MENU: General Options

ref parmname    parameter description                value  index
1.  SPECIFY     Plot characters, lines, and colors?  N
2.  AXES        Do you want to specify Axes/Labels?  N
3.  TICKMARK    Do you want to specify tick marks?   N
4.  TITLE       Do you want to specify titles?       N
5.  FTNOTE      Do you want to specify footnotes?    N

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS>
```

Figure 7-16. General Options Menu

For each option you select, you will be prompted to enter the appropriate data. For AXES, you will enter labels for X and Y and the reverse axes. The scattergram is drawn when you are through with the appropriate menus.

7.3.3 LINREG

Option 2 on the Character Output Plot menu takes you to the LINREG menu. This routine predicts a value of the dependent variable then both the actual and the dependent predicted values are plotted against the independent variable. The output graph is an overlay of dependent vs. independent and predicted vs. independent.

When you select this option, the parameter entry shown below appears. It is followed by a discussion of each parameter.

```
MENU: Produce predicted and plot with actual values

ref parmname    parameter description                  value   index
1.  DATASET     Name of the input dataset
2.  GRAPH       Name of the output graph
3.  TAG         Tag field for the output graph         *
4.  INDVAR      X-AXIS (INDEPENDANT) variable
5.  DEPVAR      Y-AXIS (DEPENDENT) variable
6.  PAGESIZE    Number of lines of output              *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT,
BACK, END, CLEAR, EXIT
GEMS >
```

Figure 7-17. Predicted Plot Menu

1. DATASET: The name of the GEMS dataset to use as the input source.

2. GRAPH: The name of the graph which will be created by the procedure.

3. TAG: A descriptive field of up to 60 characters to further describe the graph.

4. INDVAR: The name of the x-axis variable (the INDEPENDENT or regressor) variable.

5. DEPVAR: The name of the y-axis (DEPENDENT or response) variable.

6. PAGESIZE: The number of lines available on your terminal for the graph. (This parameter will only appear on the menu if you have selected the character output histogram.) Defaults are available for this parameter and are set by the terminal type which you setup when you entered GEMS:

VT100 and compatibles and all TEKxxxx except TEK4014:	24 lines
TEK4014:	32 lines

LA120:	66 lines
132 ASCII terminal:	24 lines

You may accept the default which the procedure will select for you or set your own line count.

Upon completing this menu, enter NEXT to move on to the final menu. If you are in a node other than VAXTM1, GEMS will display a warning message and end the procedure. This is because the procedure uses SAS which is only available on VAXTM1. The General Options menu shown in Figure 7-13 appears.

After specifying the items which you want to alter by changing the parameter to Y, you will be prompted for the applicable data. When finished, enter NEXT to proceed to drawing the scattergram.

7.3.4 Overlay

Option 3 from the plot selection menu (for graphics output) takes you to the Overlay menu shown below. Note that the COVERLAY menu for character output is almost identical to this one, as is the procedure.

MENU: Overlay (Superimpose two or more plots)

ref	parmname	parameter description	value	index
1.	DATASET	Name of the input dataset		
2.	GRAPH	Name of the output graph		
3.	TAG	Tag field for the output graph	*	
4.	XVAR	X-Axes variable name		
5.	YVARS	Y-Axes variable names		(1)

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

Figure 7-18. Overlay Menu

1. DATASET: The name of the dataset which contains the variables you will use as input to the Scattergram operation.

2. GRAPH: The name of the output graph.

3. TAG: A descriptive field of up to 60 characters to help you identify the graph in the future.

4. XVAR: The name of the variable for the X axis.

5. YVARS: The names of the variables to use on the Y axis. You can select up to four (4) variables to be overlaid on this graph.

If you are operating in the character output mode, a sixth option, "PAGESIZE" is available. This option allows you to specify the length, in lines, for your output page. A default value, which you may assume, is automatically set for this variable according to the terminal type you specified at the beginning of this session.

The [illegible] command tells the system to validate your datasets and variables and to proceed to the next menu.

At this menu, you will select the interpolation method to be used to connect the points on the scattergram you generate. See the discussion of interpolation methods in section 7.3.2. If you selected the character output option, you will not see this menu as the only option is NONE.

```
MENU: Select one of the following INTERPOLATION METHODS

1. Leave the points unconnected                (NONE)
2. Join the points using a straight line       (JOIN)
3. Join the points using CUBIC SPLINE method   (SPLINE)


Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

Figure 7-19. Interpolation Methods Menu

The General Options menu shown in Figure 7-13 appears next. If you have previously specified character output rather than graphic output, this menu will differ slightly. The SPECIFY parameter will allow you to specify plot symbols rather than lines and colors. The FTNOTE function is not supported for character output.

For each option you select, you will be prompted to enter the appropriate data. For the AXES parameter, you must enter labels for X and Y and the reverse axes. The final step is to enter the NEXT command to actually begin processing. An example of a single plot graphic scattergram is provided in Figure 7-21.

7.4 Isopleth (ISO)

The Isopleth procedure produces two- and three-dimensional graphic contour plots. Up to 100 levels of a contour variable may be represented. Each may be drawn with a different color and line style. A pattern may also be specified to fill each contour level.

From the main GEMS Graphics Menu, you may enter either a 4 or ISO to invoke the Isopleth procedure. The following navigational menu appears.

```
MENU: Isopleth

1. Character Output Isopleth                    (CISO)
2. Graphics Output Isopleth                     (GISO)


Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

Figure 7-20. Isopleth Menu

Choose the type of isopleth you want to create. For the graphics output option, you must be on a Tektronix or Tektronix compatible terminal. When you have selected the

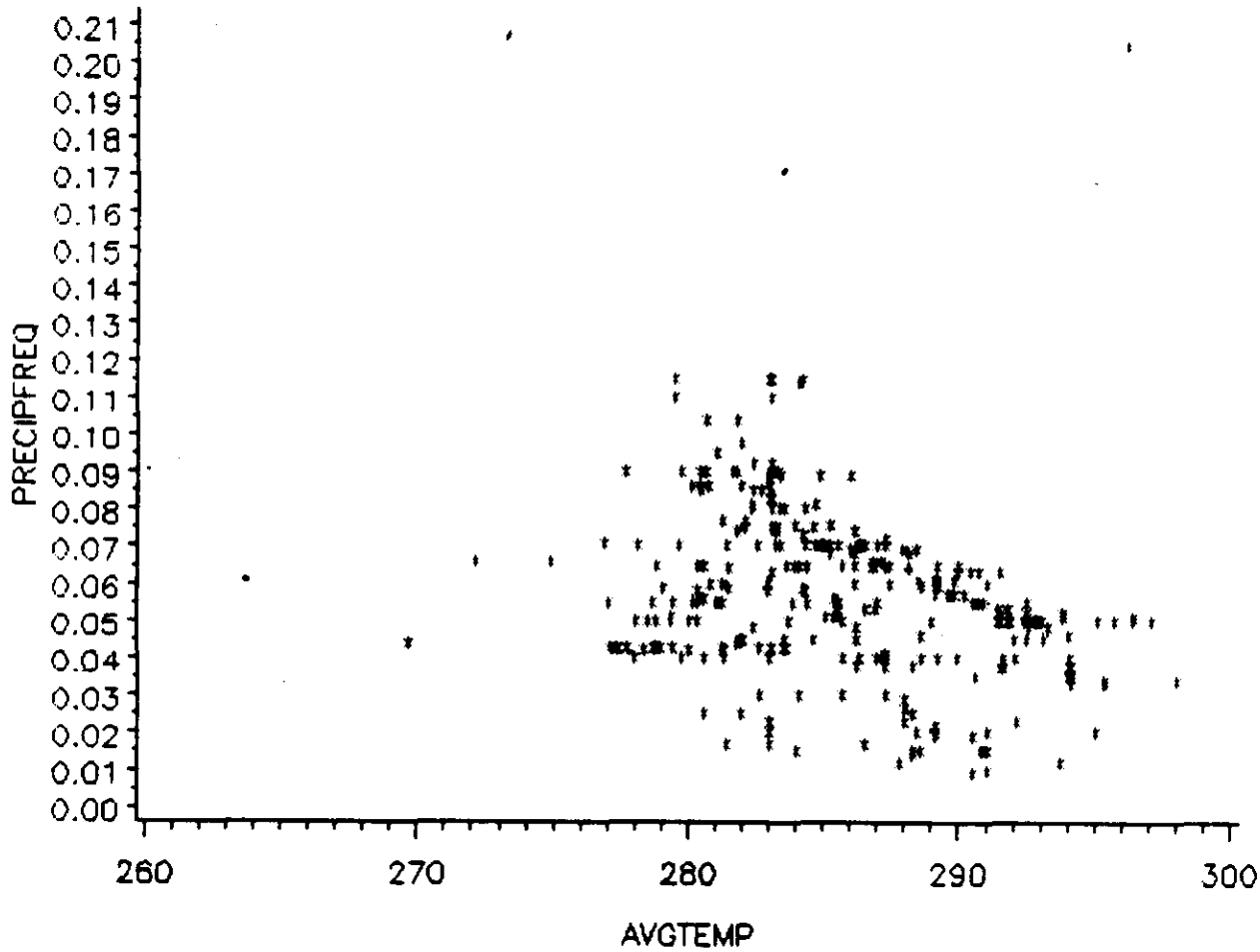

Figure 7-21. Example of Single Plot Scattergram

type of isopleth to be created, you will continue to the next menu shown below.

```
MENU: Isopleth

ref parmname   parameter description          value  index
1. DATASET     Name of the input dataset
2. GRAPH       Name of the output graph
3. TAG         Tag field of the output graph   *
4. THREED      3-Dimensional Isopleth?         N

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

Figure 7-22. Isopleth Parameters Menu

1. DATASET: The name of the input dataset.

2. GRAPH: The name of the output graph that is to be created.

3. TAG: A descriptive field of up to 60 characters to help you identify the graph in the future.

4. THREED: A yes/no switch for selecting a three-dimensional isopleth. Enter Y if you are creating a three-dimensional isopleth or N for a two-dimensional isopleth (the default is to create a two-dimensional isopleth). This option is only available for the graphic output isopleth.

After this menu, the NEXT command allows you to enter the axes variables.

```
MENU: Isopleth

ref parmname    parameter description                    value  index
1.  XVAR        X-Coordinate Variable name
2.  YVAR        Y-Coordinate Variable name
3.  ZVAR        Z-Coordinate Variable name


Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

Figure 7-23. Isopleth Axes Variables Entry Menu

1. XVAR: The name of the variable in the input dataset which is to be used for the X-coordinate values.

2. YVAR: The name of the variable in the input dataset which is to be used for the Y-coordinate values.

3. ZVAR: The name of the variable in the input dataset which is to be used for the Z-coordinate values. This parameter is needed only if you are doing a three-dimensional isopleth.

Once you have specified all of the parameters which are necessary for your graph, enter NEXT to proceed. The following prompt appears.

Do you want to interpolate using the weighted mean average (Y/N)
Default = N
GEMS >

Enter Y if you want the system to do the interpolation using an objective analysis scheme that incorporates a weighting function based on the value of the individual points. The weight for each given data point equals the reciprocal of the distance from

from that data point to the grid points where the data value is to be estimated. If you do not want this, you may either enter a *carriage return* or *N* to accept the default of NO.

The following parameter entry menu appears. This menu will be different depending on whether you selected graphics or character isopleth, and whether you are creating a two-dimensional or a three-dimensional isopleth. The appropriate menu may have only a subset of these parameters.

MENU: Options for 2-dimensional Isopleth

ref	parmname	parameter description	value	index
1.	CONTOURS	Number of contours to draw	8	
2.	LOG	Do you want a logarithmic scale?	N	
3.	AXES	Do you want to specify axes options?	N	
4.	LEGEND	Do you want a legend?	Y	
5.	LEVELS	Do you want to specify levels?	Y	
6.	TITLE	Do you want to specify titles?	N	
7.	FTNOTE	Do you want to specify footnotes?	N	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

Figure 7-24. 2-dimensional Isopleth Options

1. CONTOURS: The number of contour lines to be drawn on the graph. Character isopleths do not allow the selection of levels. Only the number of contours with the values of the levels being assigned and given in the legend. You may specify any number from 2 to 25 contours.

2. LOG: A yes/no switch for trasformation of the isoline values. Enter Y to transform the isoline values of the graph into a LOG10 scale.

3. AXES: A yes/no switch for axes labels. Enter Y if you want to specify axes labels, color of labels and other axis options.

4. LEGEND: A yes/no switch for legend specification. Specify Y for this option if you want a legend displayed giving the range of values and the corresponding pattern. This parameter will only appear for the two-dimensional isopleth.

5. LEVELS: A yes/no switch for levels. Specify Y if you want to specify the values to be used in the isopleth mapping. These values, along with the line type and color, will be shown in the legend (if you have also set legend to Y). If N has been specified, the procedure will compute default isoline levels. This parameter will only appear for the two-dimensional isopleth.

6. TITLE: A yes/no switch for title specification. Enter Y for this option if you want to specify a title of up to two lines for the graph. If N is specified, no title will be given to the graph.

7. FTNOTE: A yes/no switch for footnotes. Set to YES to allow you to specify up to two lines of text to be printed at the bottom of the graph. If you enter N instead, footnotes will not be displayed.

When the menu is complete, enter the NEXT command to continue. Depending on the selections you made in this menu, additional parameter entry menus will appear requiring the entry of the appropriate data.

If you indicated in the previous menu that you wish to specify axes options, the following menu will be displayed:

```
MENU: Options for 2-dimensional Axes Labels

ref parmname   parameter description          value   index
1.  XLABEL     Label for X Axis
2.. YLABEL     Label for Y Axis
3.  ZLABEL     Label for Z Axis
4.  CLABEL     Color of labels                WHITE
5.  CAXES      Color of Axes                  RED

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

Figure 7-25. Options for 2-Dimensional Axes Labels

The variables are self-explanatory. This menu allows you to specify the labels which will be used when drawing the isopleth. It also allows you to select the colors to be used for the labels and axes. Note that you can only specify colors on Tektronix 41xx series terminals or on a PC running Tektronix 41xx terminal emulation software. The color specification parameter will be ignored for all other terminal types. To continue from this menu to the next step, enter NEXT

If you have specified LEVELS (available only for 2-dimensional isopleth), you will be prompted to enter numbers to specify the ranges which will be used in defining the levels. You will be prompted to enter values within specified ranges. In order to complete the process, enter a carriage return for one of the values. The procedure will then execute and create the isopleth graph in your GEMS User Library. An example of a three-dimensional isopleth graph is shown in Figure 7-26.

7.5 Rose (ROSE)

The Rose procedure produces a circular graphic representation of the distribution and speed of wind patterns. It is divided into 16 clockwise-directional sectors with north as degree zero. To invoke this procedure, you may enter either a 5 or ROSE at the graphics menu. The menu shown below appears.

EXAMPLE OF THREE-DIMENSIONAL ISOPLETH

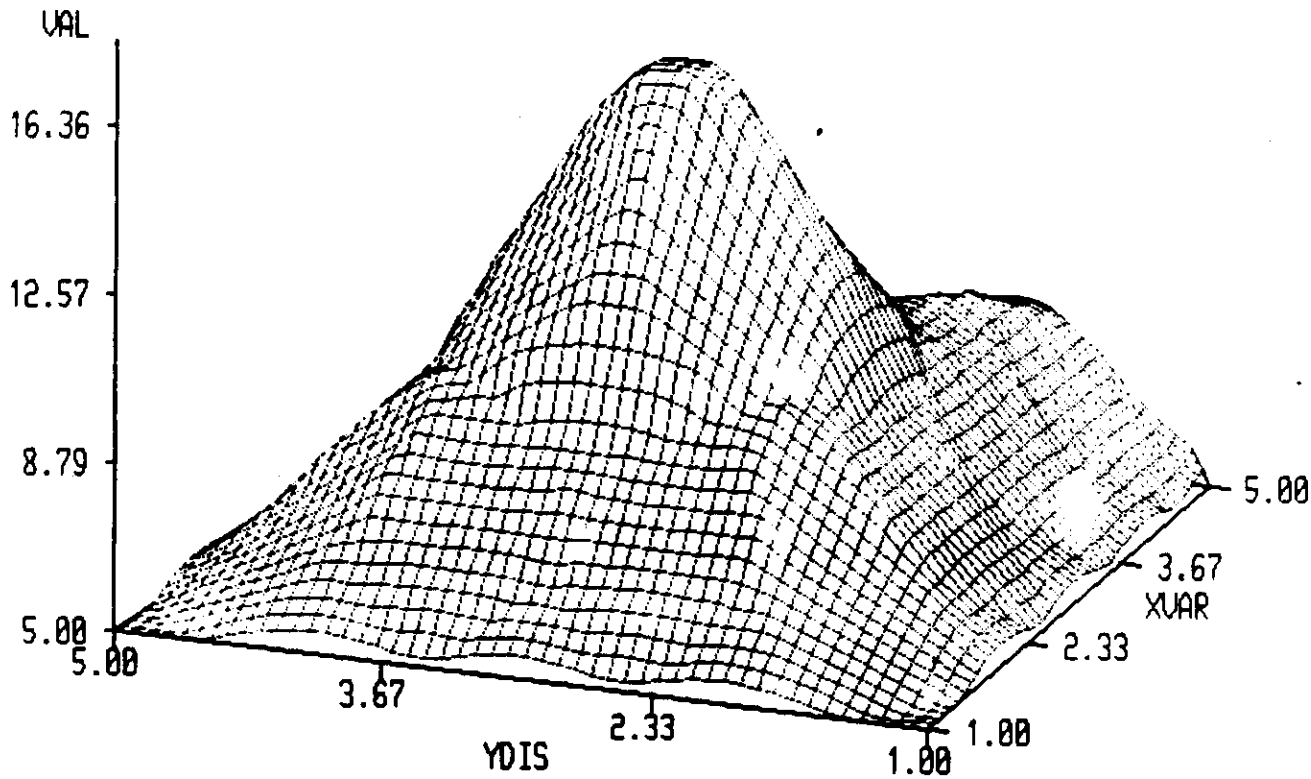

Figure 7-26. Example of isopleth

```
MENU: Select Method to Locate the STAR Station

1. Zip code
2. UTM Coordinates
3. Latitude/Longitude Coordinates (decimal degrees)
4. Latitude/Longitude Coordinates (degrees, minutes, seconds)
5. Manual input

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

Figure 7-27. Rose Methods Menu

Of the menu options, the first four are location options and are very similar. They all use the STAR database and display identical menus except for location identifier parameters. The STAR database contains the data and locations of first-order weather stations for the National Weather Service. The data available for each station consists of the STability ARray (STAR) frequency distributions characterizing the long term statistical distribution of wind speed and direction by atmospheric distribution class. The Rose procedure is performed using summary data from a user-selected specific weather station. The manual input production option allows you to directly input the wind data at the time of execution. Since the first four options are functionally the same during subsequent specification steps, we will only discuss the Latitude/Longitude Coordinate (decimal degrees) and Manual input options.

7.5.1 Produce Rose Using Latitude/Longitude Coordinates (Decimal Degrees)

This procedure displays a wind rose of the data from one of the STAR Stations in the database which GEMS maintains. The location information which you enter is used to search the database to find the nearest STAR Stations. These will then be displayed in a menu where you may select the station whose geographical profile most closely resembles the location that you entered. You may then create the wind rose of the selected portion of the data. If you select this option, the parameter entry menu shown below appears.

```
MENU: Locate STAR station by latitude/longitude (decimal degrees)

ref parmname   parameter description              value  index
1.  LAT        Latitude to locate stations         0
2.  LONG       Longitude to locate stations        0
3.  TYPE       Wind (D)irection, (S)peed, or (B)oth D
4.  ORIGIN     Origin of ROSE: (Z)ero or (M)inimum  Z
5.  GRAPH      Name of the output graph
6.  TAG        Tag field for the output graph      *

Enter one or more combinations of: reference or parameter name and
value(s) [ref1 value1, ref2 value2,...]'or a command: HELP, NEXT, BACK,
END, CLEAR, EXIT
GEMS >
```

Figure 7-28. Locate STAR Station by Latitude/Longitude

1. LAT: The latitude of your location.

2. LONG: The longitude of your location.

3. TYPE: The data which is to be displayed from that retrieved from the STAR station data. Your choices are wind direction (D), the wind speed (S) or both (B).

4. ORIGIN: The parameter used to specify whether the origin of the rose (0,0) is to be located at the point where the data is zero or the point where the data is at its minimum value.

5. GRAPH: The name of the graph which will be created in your GEMS User Library.

6. TAG: A descriptive field which can be used to further identify the contents of the graph which you create.

When you are entering the above information, remember that the GEOLIST procedure provides help in identifying zip codes or latitude/longitude coordinates. See Chapter 4 for more on the Search for U.S. Cities function.

After specifying the parameters above, a list of STAR stations will be displayed and you will then be asked to enter the index number of the station you are selecting. You will then be prompted as to whether or not you want to view the STAR summary and auxiliary data. Finally, you will be asked whether you want to perform the ROSE on selected STAR summary data. The ROSE is then produced on your screen.

7.5.2 Produce ROSE with Manual Input (ROSEMAN)

This option allows you to draw a rose of data which you supply. This can be useful if you have data for a STAR station not contained in the database which GEMS maintains, or you have other data which is useful to display in this manner. The parameter entry menu below appears if you select this option.

MENU: Produce ROSE using manual input

ref	parmname	parameter description	value	index
1.	TITLE1	First title of the graph	*	
2.	TITLE2	Second title of the graph	*	
3.	ORIGIN	Origin of ROSE: (Z)ero or (M)inimum	Z	
4.	GRAPH	Name of the output graph	*	
5.	TAG	Tag field for the output graph		

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

Figure 7-29. Rose Manual Input

1. TITLE1: The first line of the title. It will help you identify the graph. It is limited to 60 characters.

2. TITLE2: The second title line to expand the definition of the graph.

3. ORIGIN: Specify whether the origin of the Rose (0,0) is to be at the point where the data

is zero or the point where the data is at its minimum value.

4. GRAPH: The name of the output graph.

5. TAG: A descriptive field which can be used to further identify the contents of the graph.

Remember that GEMS does not support entry of more than 60 characters per line. The use of the TITLE2 variable expands this limit to 120 characters for the title. When this menu is complete, enter NEXT to proceed. The following two menus appear. You must enter all the values in one "page" before proceeding to the next "page".

MENU: Data values by sector Page 1 of 2

ref	parmname	parameter description	value	Index
1.	N	North	0	
2.	NNE	North-North-East	0	
3.	NE	North-East	0	
4.	ENE	East-North-East	0	
5.	E	East	0	
6.	ESE	East-South-East	0	
7.	SE	South-East	0	
8.	SSE	South-South-East	0	
9.	S	South	0	
10	SSW	South-South-West	0	
11.	WSW	West-South-West	0	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

Figure 7-30. Data Values for Manual Input Rose-Page 1

MENU: Data values by sector Page 2 of 2

ref	parmname	parameter description	value	index
1.	W	West	0	
2.	WNW	West-North-West	0	
3.	NW	North-West	0	
4.	NNW	North-North-West	0	

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

Figure 7-31. Data Value for Manual Input Rose-Page 2

When you have completed both "pages", enter NEXT and the rose will then be drawn on your screen. An example of a rose is shown in Figure 7-32.

STAR STATION NEW ORLEANS/MOISA LA 1960-1964

SECTOR	(FREQUENCY)
N	7.693E-02
NNE	7.893E-02
NE	9.642E-02
ENE	9.300E-02
E	5.150E-02
ESE	4.922E-02
SE	5.521E-02
SSE	6.966E-02
S	6.203E-02
SSW	5.642E-02
SW	6.292E-02
WSW	5.114E-02
W	5.554E-02
WNW	4.339E-02
NW	3.330E-02
NNW	6.442E-02

PLOT TYPE = WIND DIRECTION

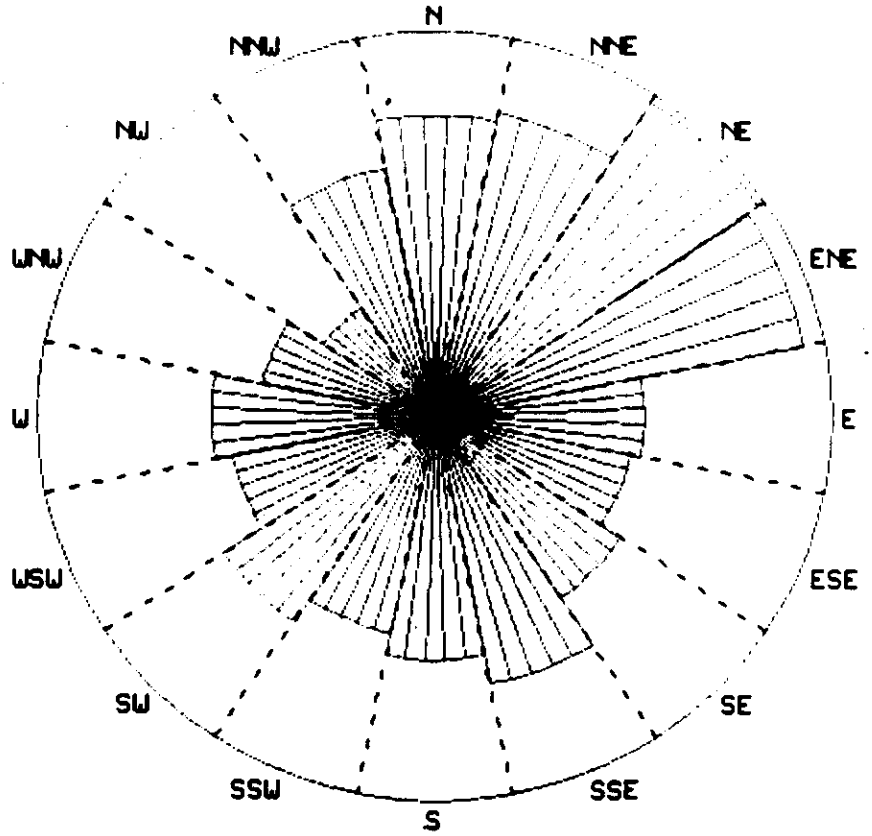

Figure 7-32. Example of Rose

Chapter 8 - Utilities

The GEMS UTILITIES procedure gives you access to a number of useful procedures, communication with the GEMS System Manager, and access to a number of useful VAX/VMS utilities and other functions which reside outside of GEMS.

```
MENU: Utilities

1. Report GEMS Problems                          (PR)
2. Invoke Statistical Analysis System            (SAS)
3. Convert an ASCII File to a SAS File           (ASC2SAS)
4. Invoke ARCINFO                                (ARCINFO)
5. Check CPU USAGE                               (TOPCPU)
6. Directory Management Utilities                (DIRMAN)
7. VAX/VMS Utilities                             (VMSUT)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 8-1. Utilities Menu

8.1 Report GEMS Problems (PR)

The PROBLEM REPORT allows you to communicate directly with the developers of GEMS, to report any problems you encounter, and, just as importantly, pass along any comments about system use. Your suggestions on system operation are often the inception of new system features or enhancements.

When you select this option, GEMS will take you into the VMS EDT editor. You will see the following:

During the editing, press CTRL-Z when finished or CTRL-C to cancel the report
*

At the editor prompt (*), you can use either the editor line mode, or if you are on a VT100 or compatible terminal, the editor video mode. To enter the line mode just enter *I* at the prompt, then type your problem report. When you are finished, enter *[CTRL-Z]* to return to the editor prompt.

At a VT100 or compatible terminal, you can use the video mode if desired. To enter video mode, enter **C** at the prompt. You can now enter your message. You can move around within your message using the arrow keys to change portions until you are satisfied with them. When you have finished your message, just enter *[CTRL-Z]* to return to the editor prompt.

Enter **EXIT** to leave the editor. When you submit a problem report, the report is automatically sent to the GEMS system manager. A copy of the report is also sent back to you as a record. The GEMS system manager will get back to you within a day or two to advise you of the problem's resolution.

8.2 Invoking the Statistical Analysis System (SAS)

The power of the Statistical Analysis System (SAS) is available when you enter the command **SAS**. For more on the use of SAS, please consult the *SAS Users Guide*. SAS is available only on VAXA (VAXTM1).

8.3 Convert ASCII to SAS Files (ASC2SAS)

To use ASCII files with SAS, you must first convert them to the SAS file format. The ASC2SAS procedure allows you to create a new SAS formatted file from an existing ASCII file. The ASCII file does not have to have a fixed format for the field positions, but it does require a unique delimiter be used to separate the fields. When you use the ASC2SAS procedure, the parameter entry menu shown below allows you to specify conversion parameters.

MENU: Convert an ASCII File to a SAS File

ref	parmname	parameter description	value	index
1.	ASCFILE	Name of the Input ASCII File		
2.	VARIABLE	Names of All Data Fields		1
3.	DELIMIT	Data Field Delimiter	*	
4.	SASFILE	Name of the Output SAS File		

Enter one or more combinations of: reference or parameter name and value(s) [ref1 value1, ref2 value2,...] or a command: HELP, NEXT, BACK, END, CLEAR, EXIT
GEMS >

FIGURE 8-2. SAS File Conversion Menu

1. ASCFILE: Name of the ASCII file to be converted to a SAS file.

2. VARIABLE: An array of variable names to use in the SAS file to represent the data fields in the ASCII file.

3. DELIMIT: Character to use as the delimiter to separate the fields in the ASCII file. Remember this character must not be used in the file for any purpose other than as the field delimiter.

4. SASFILE: Name to use for the new SAS format file you are creating. This file name may be

up to eight characters. You need not supply an extension; SAS automatically assigns an extension of .SSD.

When you have completed the parameter selection screen, enter *NEXT* to proceed. Once the file has been created, you see:

SAS file DBA2:[GEMSUSER]FILENAME.SSD
successfully created
Press RETURN for menu...

Press *carriage return* to see the UTILITIES menu.

8.4 Invoke ARCINFO (ARCINFO)

This option provides access to the ARCINFO Geographic Information System (GIS) from within GEMS. For more information on ARCINFO, see the ARCINFO documentation set. ARCINFO is available only on VAXB (CASTOR).

8.5 Check CPU Usage (TOPCPU)

TOPCPU allows you to quickly check which processes are using the most CPU time at any moment. The display will be in percent (%) of CPU time available. This allows you to check on a batch job you have submitted, or see if there are several large jobs running which are slowing down the system. You terminate the TOPCPU procedure with *[CNTRL-Y]*. (If you use [CNTRL-C] to terminate the procedure, you will have to use the EXIT command at the MONITOR prompt.)

8.6 Directory Management Utilities (DIRMAN)

The DIRMAN procedure gives you access to several useful utilities for managing files used by SFILES procedure . In order to call up this procedure, enter either *DIRMAN* or *d*.

```
MENU: Directory Management Utilities

1. Get SFILES Structure Files                    (GETSFILES)
2. Delete SFILES Structure Files                 (DELSFILES)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 8-3. Directory Management Utilities Menu

8.6.1 Get SFILES Structure Files (GETSFILES)

This procedure will present you with a copy of the files in which SFILES stores chemical structures you have entered. These files will be created in your main directory (the one you log onto initially). You should leave them there if you are going to continue to use SFILES, a procedure used to enter chemical structures for property estimation. See Chapter 2 for information about SFILES.

8.6.2 Delete SFILES Structure Files (DELSFILES)

If you want to delete your SFILES Structure Files, use this option and you will no longer have access to the structures you have stored. If you want to use them in the future, you will have to use GETSFILES again.

8.7 VAX/VMS Utilities (VMSUT)

This option selects the VMS UTILITIES menu, allowing you to access specific VMS commands from within GEMS.

The commands available from this menu have been selected based on their potential usefulness to GEMS users. If you find that you frequently exit GEMS to perform certain VMS operations, then let us know so that we can consider adding them to this menu. Use the PR procedure to tell us which functions you would like to see added.

```
MENU: VMS Utilities

1. Show Disk Quota                    (QUOTA)
2. List Directory                     (VMSDIR)
3. Edit a VMS File                    (VMSEDIT)
4. Access to VMS Mail                 (VMSMAIL)
5. Purge VMS Files                    (VMSPURGE)
6. Delete VMS Files                   (VMSDELETE)

Enter an option number or a procedure name (in parentheses) or a
command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS >
```

FIGURE 8-4. VMS Utilities Menu

8.7.1 Show Disk Quota (QUOTA)

Before you begin using a GEMS operation which will consume a large amount of disk space, you will want to first find out how much space is available. The QUOTA procedure displays your current logged user directory, your disk quota (the amount of space you are allowed to use on the disk), and the amount of space left on the disk that you are using. The space left on the disk drive is as important as the amount of space that you are allowed to use. If the disk has less space available than your quota allows you to use, you will still be limited to the space which is remaining on the disk. You should take care never to use up the remaining space on a disk.

8.7.2 List Directory

VMSDIR allows you to list the contents of a directory. The default is to list the files in your current directory, otherwise you must specify the file directory. GEMS responds to your selection of this option with the following prompt:

Enter the file directory
Default = *
GEMS >

If you want to see a listing of all files in your current directory, just enter *carriage return*, or you can specify the file for which to search and/or an alternate directory to be listed. Please see *A GEMS Users Guide to VAX/VMS* for how to specify VAX directories and file names.

8.7.3 Edit a VMS File (VMSEDIT)

This procedure allows you to access the VMS editor EDT. You will be prompted for the name of the file which you want to edit.

Enter a file name
GEMS > *{insert name of file to edit}*

The file name can be a file which is currently in existence or a new file that you are creating. For information on how to use EDT, see the appendix, *A GEMS User's Guide to VAX/VMS*.

8.7.4 Access VMS Mail (VMSMAIL)

This procedure allows you to send and receive mail through the VMS Mail Utility without having to leave GEMS. For additional information on using MAIL, see the *NCC VAX User's Guide*.

8.7.5 Purge VMS files (VMSPURGE)

Choosing the VMSPURGE procedure allows you to purge files from your VAX directories. VMS allows the creation of multiple files with the same name. These files are differentiated by version numbers as part of the name. Purging deletes all files except for the most recent version. When you select this option you will be asked to:

Enter file names to purge
Default = *
GEMS >

Enter names, separated by commas, of the files to be purged. If you accept the default (*), all files in the directory will be purged. The system will tell you which files have been deleted by the purge. When the purge has been completed, you see:

Press RETURN for menu...

Press carriage return and the VMS UTILITIES menu appears.

8.7.6 Delete VMS Files (VMSDELETE)

This option allows you to delete unwanted files from your VAX directories. If you require additional disk space to perform a GEMS function, you can use this option to delete unwanted files and free up space for your operation. You will see:

Enter file names to be deleted
GEMS >

As each file is deleted, the system will inform you with the file name. When the delete operation, has been completed you see:

Press RETURN for menu...

Press carriage return to return to the VMS UTILITIES menu.

Chapter 9 - Information and News

The Information and News operation is the newest function available in GEMS. It allows the GEMS support staff to get more information about GEMS to you, the GEMS user. The procedures currently available under this operation allow you to communicate with the GEMS staff (FEEDBACK) and allow the GEMS staff to post information on the modifications and updates to GEMS which will be of interest (NEWS and BULLETINS).

The Information and News menu appears if you choose this option.

```
MENU: GEMS Information and News

1. Send Messages to GEMS Staff                (FEEDBACK)
2. Type GEMS News Message                     (NEWS)
3. GEMS Bulletins                             (BULLETINS)

Enter an option number or a procedure name (in parentheses)
or a command: HELP, HELP option, BACK, CLEAR, EXIT, TUTOR
GEMS
```

FIGURE 9-1. Main Information and News Menu

Each option is explained below.

9.1 Send Messages to GEMS Staff (FEEDBACK)

The FEEDBACK procedure allows you to communicate directly with the developers of GEMS to pass along

suggestions. Your suggestions are important in the development of additional capabilities for the system.

When you select this option, GEMS places you into the VMS EDT editor for entry of your message. As you enter your message, you will see the following:

During the editing, press CTRL-Z when finished or CTRL-C to cancel your message
*

At the editor prompt (*), you can use either the editor line mode, or, if you are on a VT100 or compatable terminal, the editors' video mode. To enter text in line mode, enter L at the prompt and then type your message. When you have finished, enter CTRL-Z to return to the editor prompt.

At a VT100 or compatable terminal, you can use the video (or full screen) mode if desired. To enter video mode, you should enter G at the prompt. You can now enter your message. You can move around within your message by using the arrow keys to change portions until you are satisified with them. When you have finished your message, just enter CTRL-Z to return to the editor's prompt.

Enter EXIT to leave the editor. When you exit the message will automatically be sent to the GEMS system manager. A copy of the message is also sent to you as a record.

9.2 Type GEMS News Message (NEWS)

The NEWS procedure retypes the GEMS Welcome message. This is the message which is displayed when you run GEMS. This message will normally contain a list of bulletins which are available under the BULLETINS menu. You should be sure to check the news message for new bulletins each time you use GEMS.

9.3 GEMS Bulletins (BULLETINS)

The GEMS News Bulletins give you information on various GEMS procedures being added to the system or updated. The News Bulletins are listed on the GEMS Welcome Message as you log into the system. Please check out any new bulletins as they are added to the list. These can give you some very useful information on the modifications and updates to GEMS.

This Page Blank Intentionally

Chapter 10 - Datasets

Use of GEMS' chemical fate and exposure models requires a variety of input information including population and demographic information, information on industrial discharges, chemical properties data, and cartographic data. Much of this data can be found in the datasets available in GEMS.

The databases in GEMS consist of two distinct types of files. Those installed in the GEMS File Management System are available to GEMS users at all times. This first type of data is subject to GEMS data manipulation, analysis, and display capabilities. These files are described in Section 10.1 of this chapter. The second type of data files are typically large and/or specialized files which may be specific to some models. This type of data is accessible only by special retrieval capabilities (e.g., during a model run) not present in the GEMS File Management System. These files are described in Section 10.2 of this chapter.

10.1 GEMS Datasets

Those databases to which you have access through the GEMS File Management System are discussed in this section. Each is presented along with certain specifications about each, including:

- Dataset name
- Status of the dataset (on-line or off-line)
- Full title of dataset

These files have been grouped by the type of information they contain (e.g., environmental characteristics or chemical release sites). A few of these datasets are automatically accessed by GEMS models during the model run. In most cases, however, the information must be accessed by the user separately.

10.1.1 Datasets to Support Environmental Assessments

10.1.1.1 Population Density and Exposure Estimates

COUNTYPOP

DATASET NAME	STATUS	TITLE
COUNTYPOP	ON-LINE	COUNTY POPULATION FIGURES SUMMED FROM MARF C8083 FILES

The COUNTYPOP dataset was built by processing the Master Area Reference File (MARF). The populations by county were added for the various populations in MARF. See the description below of MARF for more information.

1980 MASTER AREA REFERENCE FILE (MARF)

DATASET NAME	STATUS	TITLE
C8083ST	ON-LINE	MARF 80 UPDATED 1983 POPULATION BY AGE, SEX, AND HOUSEHOLD INCOME DISTRIBUTION

The Master Area Reference File of the 1980 Census of Population and Housing was prepared by the Bureau of the Census. The file has a variety of location identification and population information, including region, state, county, place, census tracts, enumeration districts or block groups, population count by race, number of occupied and owner-occupied housing units, group quarters, and number of families for all the enumeration districts/block groups for the continental United States, Hawaii, and Alaska.

The complete MARF 1980 Census file is your first source to identify household and population by racial groups at any required geographic level. County aggregate populations have been created from this file.

The 1983 Census Update file of demographics data was added to existing records of the MARF 1980 Census data in order to provide complete specification of geographic location for block groups and enumeration districts (BG/EDs). It is a proprietary product of Donnelly Marketing, Inc., and is available only to EPA users and to contractors engaged in EPA projects. The database includes the following demographic estimates for 1983:

- Total population
- White population
- Black population
- Spanish population
- Other population
- Household population
- Group quarter population
- Numbering household
- Average household
- Household income distribution
- Population by sex and age

The data have been organized by state. The dataset name is C8083ST, where ST is replaced by the two-letter United States Postal Service (USPS) state abbreviation.

A subset of the 1983 Census Update file is accessed by the atmospheric models to calculate the population exposure to various levels of the chemical in question, the cumulative population exposed within a specific distance and/or direction of the source of release, and the cumulative population exposure.

10.1.1.2 Atmospheric Assessments

STAR

DATASET NAME	STATUS	TITLE
STAR	ON-LINE	STability ARray

The meteorological data for this dataset, except mixing height data were obtained from the National Climatic Center in Asheville, North Carolina. The mixing heights (annual average) for the contiguous United States were obtained from a tabulation of morning and afternoon heights provided by the Oak Ridge National Laboratory. A half-degree cell resolution was used.

The STAR data files in GEMS contain meteorological data for 394 first-order weather stations in the continental United States. STAR data consists of frequencies of wind direction by wind speed classes for up to seven atmospheric stability categories. Wind directions are given for sixteen sectors corresponding to compass points. There are six wind-speed classes. The data are available as annual average frequencies based on a minimum of five consecutive years of data.

In addition, the database contains the annual average afternoon and nocturnal mixing height, indices of the stabilities, annual average, maximum and minimum ambient air temperature, fraction of time it rains, and average intensity of rainfall.

These data are also used automatically as input for climatological atmospheric models like ISCLT and TOXBOX. The STAR data are accessible for display using the ROSE graphics procedure in GEMS.

10.1.1.3 Soil and Groundwater Assessments

SOILS-5 and MUUF

DATASET NAME	STATUS	TITLE
SOILS-5	ON-LINE	THE SOILS-5 DATABASE
MUUF	ON-LINE	MAP UNIT USE FILE

The SOILS-5 database is a U.S. Department of Agriculture (USDA) Soil Conservation Service database. It contains, among other things, soil property data from nationwide surveys. It is used by various soil and multi-media models in GEMS. Data for more than 28,000 soil units are contained in the database. Each soil unit has a unique SOILS-5 record number.

The data are organized into four groups. The first group contains those soil parameters which have only a single value per soil unit. The second group contains the layer-dependent data (i.e., soil parameters which have a value for each soil layer in the soil unit). The third group contains a general description of the soil unit. The fourth group contains a descriptive modifier for the soil unit identified by a modifier code. The link between the first three groups is the SOILS-5 record number. The fourth group is linked to the first by the descriptive modifier code.

The Map Unit Use File (MUUF) contains location data for the soil units in the SOILS-5 database. Data are available for 2571 counties throughout the U.S. The data are organized by map units, identified by a map symbol, within counties. Each map unit may consist of from one to three map unit components (soil units), as identified by the SOILS-5 record number. Each record in the file is a soil unit (map unit component). The dataset parameters include the areal coverage of each map unit, land slope data, and layer depths which are used to modify the layer depths in the SOILS-5 database. The link between MUUF and the SOILS-5 database is the SOILS-5 record number.

CLIMATE DATA FILE

DATASET NAME	STATUS	TITLE
CLIMATE	ON-LINE	NATIONAL CLIMATIC DATA BASE

The Climate database was created by the National Climatic Data Center, National Oceanic and Atmospheric Administration (NOAA) in April of 1986. It can be used to provide input data for three models currently implemented in GEMS: SESOIL, TOXSCREEN, and MICROBE- SCREEN. The database consists of two separate groups of variables.

The first group consists of climate station identification information for 3,225 climatic stations throughout the United States. Also included in this group are county names and county FIPS codes for each station. The station identification number is a six-digit code that is composed of the two-digit state FIPS code and the four-digit station number.

The second group is a combination of two types of statistics:

- 26 monthly statistical parameters for each climate stations. For each parameter, three different statistics of monthly data are given: 1) weighing gauge, 2) Fisher Porter gauge, and 3) combined gauge statistics.

- 4 monthly statistics related to evapotranspiration on a state climatic division basis. The four statistics are mean monthly wind speed (mph), mean monthly relative humidity (%), mean monthly cloud cover (fraction), and mean monthly temperature (^{o}F).

Although the four evaporation related statistics are available on a climatic division basis, they were applied to each climatic station along with the precipitation statistics. Also, an additional variable, ALBEDO, which represents the monthly short-wave albedo of the land surface, was incorporated into the database, since this variable is required by SESOIL and other related models. The albedo data were estimated from published values for 70 specific locations across the U.S. For each of the 3,225 stations, the closest site having albedo data

(based on the latitude/longitude coordinates and radial distance) was incorporated into the database.

For more information on the proper use of this data, consult *Implementation of the Climatic Data Base in GEMS* (GSC-TR8755, 1987), prepared by GSC.

DISPERSE

DATASET NAME	STATUS	TITLE
DISPERSE	ON-LINE	DISPERSIVITY AND OTHER AQUIFER DATA

Dispersivity and other aquifer data were obtained from 22 site specific studies. 18 of the studies consisted of dispersivity values obtained by model calibration, and 4 were based on two-well tracer tests. Select data from these tests were used in the creation of the "DISPERSE" dataset.

This data is of particular usefulness in the modeling for groundwater dispersivity in such GEMS models as AT123D and SWIP. For more information on the use of this dataset, consult *Groundwater Modeling Scenarios for Assessment Applications: Part II - Groundwater Dispersion Information from Case Studies* (GSC- TR8641, 1986).

VULNERGW AND DRASTIC

DATASET NAME	STATUS	TITLE
VULNERGW	ON-LINE	GROUNDWATER VULNERABILITY SCORES BY COUNTY FROM DRASTIC FACTORS (3144 RECORDS, 31 VARIABLES)
DRASTIC	ON-LINE	SEVEN GROUND WATER POLLUTION FACTORS BY COUNTY (3144 RECORDS, 58 VARIABLES)

EPA's Office of Drinking Water working jointly with the Office of Pesticide Programs have developed this information from the Groundwater Vulnerability Assessment Study.

EPA scores every county in the U.S. to indicate their groundwater vulnerability. A modified DRASTIC numerical system evaluation of the vulnerability of aquifers to pollution beneath a land area determines the score. DRASTIC is an acronym for the seven factors considered to be most significant in affecting the groundwater pollution potential. DRASTIC uses the groundwater regions defined by heath. The factors are:

- (D)epth of water
- (R)echarge net
  - (A)quifier media
  - (S)oil media
  - (T)opography slope
  - (I)mpact of vadose zone
  - (C)onductivity of the aquifer

The land area of a county may have areas which are rated differently for each of the DRASTIC factors. The DRASTIC dataset has the ratings of the seven DRASTIC factors for the land area of the county and the percentage of the land area covered by a particular rating.

The ratings or weights of the seven factors are accumulated in three different ways to provide a composite score. The Heath groundwater regions represented in the county and their hydrogeologic settings along with the corresponding percentages are shown in the dataset.

The VULNERGW dataset may be used graphically and analytically to classify an area of a county according to the potential for pesticide contamination of groundwater. The DRASTIC dataset allows you to do the same analysis with any of the individual factors which are significant in aquifer pollution.

GROUNDWATER PARAMETER FILE

DATASET NAME	STATUS	TITLE
GRNDWAT	ON-LINE	GROUNDWATER DATA FOR U.S. CITIES

Versar, Inc. compiled the information stored in this dataset for OTS. Measured data were gathered for fourteen cities in the contiguous United States. The measured data was obtained from a variety of sources, including state geological surveys, state water resources departments, and the USGS. No single source provided information on every requisite model parameter, so some estimated data were combined with the limited measured data to provide complete unit records for those locations having measured data.

This dataset was developed specifically to meet the needs of the AT123D groundwater model. The dataset contains the following groundwater parameters:

- Aquifer thickness
- Porosity of the groundwater zone
- Storage coefficient
- Seepage velocity
- Hydraulic conductivity
- Hydraulic gradient
- Transmissivity
- Soil density

TANKSITES

DATASET NAME	STATUS	TITLE
TANKSITES	ON-LINE	LOCATION, CLIMATE, SOIL AND GROUNDWATER IDENTIFIERS FOR SPECIFIC SITES

GSC developed this dataset in support of the Leaking Underground Storage Tank (LUST) survey for EPA. From this survey, specified tank sites were identified, and data for these sites were extracted from the SCS (Soil Conservation Service) published county soil surveys, SOILS-5 and Map Unit Use File (MUUF) databases from SCS, USGS NAWDEX Groundwater

Site Inventory (GWSI), USGS reports, and National Weather Service (NWS) data tapes.

The TANKSITES file contains four types of data:

- Site location and identification
- Climate
- Soil
- Groundwater/geology

The site location and identification data includes location identifiers that aid in determining the number of sites within a particular area and the location of the site on a USGS topographic map. These location identifiers include LUST and PSU (Primary Sampling Unit) regions; state and county FIPS codes; HU codes; specific site locators including site ID, latitude, longitude, and approximate elevation; USGS topographic quadrangle information providing easy geographic location including the topographic quadrangle map name, the map scale; and Soil Survey Area information that provides the Soil Conservation Service County Soil Survey name, the year the survey was published, and the survey area code.

Climate data are taken from the closest applicable weather station according to the NWS State Climatic Divisions (SCDs) retrieved from the Geoecology database. Soil data includes information which may be useful for prediction of possible tank leaks such as the seasonal high water table, the availability of C-Horizon (subsoil) information, and the relative corrosivity to steel and concrete. The data also includes Soil Map Unit numbers that may be used to find a SOILS-5 record number from Map Unit Use File (MUUF) to retrieve additional information such as permeability, pH, and percent clay.

The groundwater/geology data includes heath ranges for transmissivity, hydraulic conductivity, and porosity for the major groundwater regions, groundwater and geologic descriptions from USGS publications, and water and well information from NAWDEX GWSI such as site resolution, well usage description, and depth to groundwater.

You will find TANKSITES useful in many modeling studies. The general description of TANKSITES may be augmented with more specific information about soil, climate, and groundwater from other datasets.

10.1.1.4 Surface Water Assessments

FRDSPWS

DATASET NAME	STATUS	TITLE
FRDSPWS	ON-LINE	FEDERAL REPORTING DATA SYSTEMS PUBLIC WATER SYSTEMS

The EPA Office of Drinking Water's Federal Reporting Data System (FRDS) is the source of data for this file. This file is a subset from the dataset resident on the EPA IBM mainframe.

The FRDSPWS dataset contains information on about 60,000 public water systems (PWS). As many as fifteen different water sources supply a single PWS, and thirteen different types of treatment are provided for the water.

To facilitate the organization of this much information, the dataset has been divided into two groups: 1) general PWS information, and 2) PWS water source and treatment information. PWS water source information may be cross referenced through the common data variable, PWS ID. Variables within a group retain their group identity whenever accessed by GEMS routing. The general PWS information group contains two types of data: facility information and facility monitoring information. Facility information consists of PWS location and service area identifiers such as EPA region code, PWS ID, PWS type, treatment class code, service area type, PWS mailing address, STFIPS, zip code, population of service area, and other relevant information. The facility monitoring information includes capacity of PWS design, starting and ending dates for the monitoring of bacteriological, organic and radioactive contamination, maximum daily air temperature, whether or not the PWS carries out its own sampling, requirement and frequency of other types of sampling (nitrate, etc.), and other relevant information.

The PWS water source and treatment information includes name of the water source, type of source, source availability, river basin code, location of the water source (latitude/ longitude) and various types of treatment for the water such as sedimentation and coagulation.

If your modeling study needs to know general information about PWSs, this dataset is useful. You can always augment information in this dataset with other datasets through the use of common identification variables.

GAGE FILE

DATASET NAME	STATUS	TITLE
GAGE	ON-LINE	STREAM-FLOW RATES

If you are involved in water quality studies, waste load allocations, dilution studies, and advanced waste treatment assessments, then you will be interested in this database which has been assembled from four different sources. It's the Stream Gage Inventory File (GAGE). EPA's Office of Water Regulations and Standards gathered information from the National Water Data Exchange (NAWDEX), the Master Water Data Index (MWDI), the Basic Characteristics File (BCF), and the STORET Flow File (USGS Daily Value Flows). Cataloging units and reach numbers were determined by EPA for the USGS gaging stations identified in the Basic Characteristics File. These stations are considered to have the longest periods of natural flow data and are available in water quality assessment work. For reaches without gage stations, EPA developed a method of estimating stream flows from drainage basic characteristics.

The GAGE file is a database of stream flow rates monitored consistently by nearly 100,000 stream gaging stations from coast to coast cataloged by reach number (catalog unit number and segment number). Information stored includes location of gaging stations (in latitude and longitude coordinates), monthly flows, date of update, drainage area, calculation method, mean annual flow and seven day/ten year low flow.

WATER SUPPLY DATABASE

DATASET NAME	STATUS	TITLE
WSDB	ON-LINE	EPA WATER SUPPLY DATABASE

EPA's Office of Drinking Water and Office of Water Regulations and Standards pooled their efforts to prepare this important dataset. The STORET system provided much information on the Inventory of Public Water Supplies. Additional information on public water supplies was provided by the Public Health Service. Intake data from the FRDS Water Supply File was added along with information from various state inventories and publications such as the *Public Water Supplies of the 100 Largest Cities in the United States*, published by the USGS in 1962.

The Water Supply Database contains available surface water supply data in the U.S., including the number of surface water utilities, the total population served, the number of intake points and the number of STORET water quality stations associated with surface water utilities. The data consists of 32,223 records. Intake points are identified by latitude and longitude coordinates as well as the USGS HU code and reach number. You will find it useful to use the HU codes and reach numbers in the database facilities in other EPA files such as the Industrial Facilities Discharge File and the GAGE File to assess water quality and exposure on a basin-wide or river reach basis. The inclusion of the EPA agency station codes for water quality data in the immediate vicinity of surface water intake points also allows you to quickly identify water quality sources in drinking water quality assessment studies.

10.1.1.5 Multi Media Assessments

GEOECOLOGY DATABASE

DATASET NAME	STATUS	TITLE
AGRICULTURAL		
GEOECOA06	ARCHIVE	U.S. COUNTY 1978 CROP AREAS AND YIELDS, AG LAND USE

BASE DATA		
GEOECOB01	ARCHIVE	U.S. COUNTY CODES, NAMES AND CENTROIDS - 1970
GEOECOB02	ON-LINE	U.S. STATE NAMES AND INDICES 1970 (EPA, BBS, NOAA)
GEOECOB08 ,	ON-LINE	U.S. COUNTY GENERAL REGIONAL INDICES - 1970-75
CLIMATIC		
GEOECOC09	ON-LINE	EAST.U.S. COUNTY MONTHLY AVERAGE EVAPORATION BY COUNTY 1941-1970
GEOECOC11	ON-LINE	U.S. SCD MONTHLY AVERAGE TEMPERATURES BY DIVISION - 1941-1970
GEOECOC17	ON-LINE	U.S. COUNTY CLIMATIC DIVISION INDEX - 1970
GEOECOC20	ON-LINE	EAST U.S. COUNTY RAINFALL pH % H ON LOADING - 1974
FORESTRY		
GEOECOK04	ON-LINE	U.S. COUNTY FOREST AREA, VOLUME, AND GROWTH
GENERAL DATA		
GEOECOG01	ON-LINE	DATASET TITLE AND CHARACTERISTIC
GEOECOG02	ON-LINE	DATASET VARIABLES AND LABELS
AIR QUALITY		
GEOECOK04	ON-LINE	U.S. COUNTY AIR QUALITY NONATTAINMENT AREAS (EPA) - 1979
POPULATION		
GEOECOP01	ON-LINE	U.S. COUNTY POPULATION BY SEX AND 5 YEAR AGE CLASSES - 1970
TERRAIN		
GEOECOT03	ARCHIVE	U.S. COUNTY SOILS AT MAP UNIT LEVEL - 1967

GEOECOT07	ON-LINE	EAST U.S. SOIL ORDER, SUBORDER, MAP UNIT DIRECTORY - 1967
GEOECOT08	ON-LINE	EAST U.S. GREAT SOIL GROUPS CHARACTERISTICS - 1967
GEOECOT13	ON-LINE	U.S. COUNTY SOILS AT SOIL ORDER LEVEL - 1967
GEOEW10W9	ON-LINE	U.S. COUNTY ENDANGERED SPECIES CODES DICTIONARY (BNL FILE) - 1977

The information in these datasets was gleaned from the Geoecology database of the Environmental Science Division, Oak Ridge National Laboratory. Included are environmental data on agriculture, county base data, climate, vegetation, forestry, air quality, land, natural areas, population, water quality, terrain (soils), and wildlife (endangered species). If you are involved in the exposure assessment process, many of these will be valuable to you. For example, climatic parameters include annual average and monthly maximum and minimum temperatures and monthly precipitation and potential evapotranspiration. The agricultural files from the 1978 Census of Agriculture contain area-under- cultivation and yield statistics by county for a variety of commercially important crops. Wildlife information includes habitats for endangered species by county. Terrain data includes county soil types, together with some physical and chemical properties for general soil classes.

You can use the soil data in the Geoecology database to support the SESOIL and TOX-SCREEN models for the analysis of long-term environmental pollutants in the unsaturated soil column above the groundwater. Archived files are off-line and can only be retrieved by request. If you wish to use an archive file, use the FEEDBACK procedure in the the Information and News operation (section 9.1) to inform the GEMS support staff that you wish to de-archive the appropriate datasets.

10.1.2 Datasets Identifying Chemical Release Sites

THE INDUSTRIAL FACILITIES DISCHARGE FILE (IFD)

DATASET NAME	STATUS	TITLE
IFDDIR	ON-LINE	STREAM INFORMATION FOR DIRECT DISCHARGERS
IFDIND	ON-LINE	STREAM INFORMATION FOR INDIRECT DISCHARGERS
POTW	ON-LINE	STREAM INFORMATION FOR POTWS

The Industrial Facilities Discharge (IFD) file was assembled by the EPA Office of Water Regulations and Standards from three sources:

- National Pollutant Discharge Elimination System (NPDES) permit files located in EPA Regional Offices
- EPA's Needs Survey File
- Permit Compliance System (PCS) Database

IFD contains a comprehensive database of point source dischargers likely to discharge one or more of the 129 priority pollutants as defined in the National Resources Defense Council Settlement Agreement. There are about 88,174 facilities in the IFD file including NPDES-permitted facilities (direct dischargers plus Publicly Owned Treatment Works (POTWs)) and facilities which discharge into POTWs (indirect dischargers).

The IFD file is divided into three separate GEMS datasets:

- IFDDIR contains data on industrial direct dischargers
- IFDIND contains data on indirect dischargers
- POTW contains data on POTWs

IFDDIR contains about 28,000 records on direct dischargers (no POTWs included). It is divided into three data groups: 1) general facility information, 2) effluent guidelines information,

and 3) facility pipe data. You can use these three data groups as a single GEMS dataset. Specific facility information may be cross referenced by the following common data variables: 1) facility NPDES number, 2) state and county FIPS code, and 3) facility cataloging unit code.

The industrial facilities in the IFDDIR dataset may be linked to the REACH and GAGE files by a hydrologic unit (HU) code and the EPA stream segment number. This linkage of IFDDIR with stream information allows hydrologic pinpointing of any or all facilities along waterways, and allows you to perform hydrological analysis of point discharger information when assessing potential pollution problems downstream or upstream of a point and between two stream locations.

IFDIND contains about 12,000 records on industrial facilities that discharge waste through other facilities (usually POTWs) that are then responsible for the waste disposal. This dataset is organized into two groups, facility level and pipe level. Facility information consists of name, address, NPDES number, city, county, total facility flow, SIC codes, and the identification of the POTW receiving the indirect discharge flow. The pipe level information includes location (longitude/latitude), pipe flow, pipe SIC code, discharge type (process, cooling, etc.), receiving water, and other pipe related data. For more information on actual POTWs, read the discussion of NEEDS86 dataset.

The industrial facilities in the IFDIND dataset may be linked to the POTW dataset by the NPDES permit number. This linkage allows you to estimate the amount of industrial waste water which is treated by POTWs, and subsequently, discharged into waterways, and lets you analyze point discharger information when assessing potential pollution problems for selected streams or POTWs.

NEEDS86

DATASET NAME	STATUS	TITLE
NEEDS86	ON-LINE	SELECTED WATER TREATMENT PLANT DATA FROM 1986 NEEDS SURVEY

The NEEDS86 file is the *1986 Needs Survey Report to Congress, Assessment of Needed Publicly Owned Wastewater Treatment Facilities in the United States*. This survey was performed by the EPA Office of Municipal Pollution Control, Municipal Facilities Division, Needs and Priorities Branch.

More than 32,000 POTWs around the country have provided waste treatment facility information on nearly one hundred items such as unit treatment process, influent and effluent characteristics, flow rates, and population served. Each POTW is identified by authority/facility, zip code, state FIPS code, county FIPS code, and SMSA code.

You can use data from this file to determine both the treatment processes employed by the POTWs and the POTW waste water flow rates. This is helpful when you are estimating the degradation of the pollutants by treatment and the concentration of pollutants in the effluent.

STATESIC

DATASET NAME	STATUS	TITLE
STATESIC	ON-LINE	INDUSTRIES (SIC) AND NUMBER OF EMPLOYERS BY STATE

The STATESIC dataset has data on employment figures by state, classified by the Standard Industrial Code (SIC) for employment categories. This dataset was created by summation from the 1983 County Business Patterns File from the Bureau of the Census. See the description of the CNTYSIC dataset below for more complete information on this.

This dataset is useful if you need data on the number of employees in a particular industry. These data can be used in calculating industrial releases for an area based on the number of employees in the industry and the number of employees in the state.

CNTYSIC

DATASET NAME	STATUS	TITLE
CNTYSIC	ON-LINE	COUNTY BUSINESS PATTERNS BY SIC CODE

Taken from the 1983 County Business Patterns file from the Bureau of the Census, CNTYSIC provides data on the total number of U.S. industries or establishments and the number of employees as of 1983. The industries and employees are classified hierarchically by a 4-digit Standard Industrial Classification (SIC) Code and by county. Title 13, Section 9 of the U.S. code protects the identity of individual employers, and some of the figures were calculated from the number of establishments by type of employment. If the number of establishments by employment size were not explicitly available, the mid-point of the employment size class for data withheld to avoid disclosure was used. A flag in the database identifies the means by which the figure was derived.

EPA has used these data to estimate the geographic distribution of the use of chlorinated solvents. You may also find it useful when using the GEMS Atmospheric Modeling Subsystem (GAMS) to estimate geographic distribution of other types of emissions by county and 4 digit SIC code.

HWDMSFACILITY

DATASET NAME	STATUS	TITLE
HWDMSFACILITY	ON-LINE	HWDMS FACILITY FILE

EPA requires all generators, treaters, storers, and disposers of hazardous waste to provide information on their activities. These reports, RCRA Part B, provide the data for this file.

The dataset contains data on approximately 150,000 hazardous waste facilities. The number of records for each facility depends on the type of facility and amount of waste they process. These data are divided into two groups: 1) facility information and 2) design capacity information. Design capacity information may be cross referenced by the EPA facility ID number. Facility information includes EPA facility ID number, facility name, mailing address and location (latitude/longitude), activity code, transportation mode, owner/operator name, address, telephone number, and other information such as the state district code. The design capacity information includes design capacity process code, process amount, unit of measure, and capacity status indicator. You can use the HWDMS file to locate more specific information about such facilities for use in modeling studies. Since the file contains location variables, it may be used in conjunction with geographical mapping procedures in GEMS such as the State/County Map procedure to plot the location of hazardous waste facilities in the country.

1977 ECONOMIC CENSUS DATA

DATASET NAME	STATUS	TITLE
MNFR77.CO/PL	ARCHIVE	CENSUS OF MANUFACTURERS 1977 GEOGRAPHIC AREA SERIES BY COUNTY AND PLACE
MNFR77.SMSA	ARCHIVE	CENSUS OF MANUFACTURERS 1977 GEOGRAPHIC AREA SERIES BY SCSA/SMSA
MNFR77.STATE	ARCHIVE	CENSUS OF MANUFACTURERS 1977 GEOGRAPHIC AREA SERIES BY STATE
LOC.ST	ARCHIVE	CENSUS OF MANUFACTURERS 1977 LOCATION OF MANUFACTURING PLANS BY EMPLOYEE SIZE AND STATE
WHS77.WA8	ARCHIVE	CENSUS OF WHOLESALE TRADE 1977 BY SMSA, COUNTIES, AND CITIES

SI77A1.SMSA	ARCHIVE	CENSUS OF SELECTED SERVICE INDUSTRIES 1977 BY SMSA
COMMTR_SIC28	ARCHIVE	CENSUS OF TRANSPORTATION 1977 COMMODITY TRANSPORTATION SURVEY OF ORIGIN/DESTINATION COMMODITY FLOW (SIC 28)
COMTR_SIC29	ARCHIVE	CENSUS OF TRANSPORTATION 1977 COMMODITY TRANSPORTATION SURVEY OF ORIGIN/DESTINATION COMMODITY FLOW (SIC 29)

The 1977 Economic Census Data is a comprehensive census of U.S. industrial and business activities conducted by the Bureau of the Census. Files include:

- Census of Manufacturers, 1977
  - Geographic area series
  - Industry series
  - Location of manufacturing plants
  - Fuels and electricity consumed in 1977 and 1978
- Census of Wholesale Trade, 1977
  - Geographic area series
- Census of Selected Service Industries, 1977
  - Geographic area series
- Census of Transportation, 1977
  - Commodity transportation survey
  - Truck inventory and use survey

U.S. industry and business establishments are classified by a 2 to 6 digit Standard Industrial Classification (SIC) code. The Geographic Area Series of files provides information at various geographic levels, including state, SMSA, SCSA, county, and incorporated city. Industries are classified hierarchically by 2 to 4 digit SIC code. The information for each industry group identified by SIC code includes number of industries or establishments, number of employees and their wages, number of production workers and their hours, receipts or value added by the manufacturer for 1977, and figures for

comparison of the number of employees and value added by the manufacturer.

The Industry Series summarizes details of industry groups for the whole of the U.S. The data includes size of employment classes and product class identified by a fifth digit in a 5-digit SIC code. Files from the Industry and Geographic Series have been selectively implemented in GEMS to provide summary information for the maximum SIC digit classification at the lowest possible geographic level.

The Location of Manufacturing Plants file provides information on the number of establishments in seven employee-size categories for four-digit SIC code groups. The geographic areas include states, counties, and places within states.

The Commodity Transportation Survey provides statistics on the volume and characteristics of intercity commodity shipments by origin/destination SMSA pairs. The Truck Inventory and Use Survey provides data on the physical and operational characteristics of the nation's truck resources, including weight of cargo by type.

The Fuel and Electricity Energy Consumed files present statistics by industry groups and by industry. Statistics on quantities and costs of specified fuels and electrical energy used for heat and power are available by the 4-digit SIC code for the U.S., the 2 to 3 digit SIC code for states, and the 2-digit SIC code for SMSAs. Current files include data from the 1978 and 1979 surveys.

Archived datasets are off-line and can only be retrieved by request. If you wish to use an archive file, use the FEEDBACK procedure in the the Information and News operation (section 9.1) to inform the GEMS support staff that you wish to de-archive the appropriate datasets. You will find this information useful in exposure assessments to determine the sizes of potentially exposed occupational populations and the distribution and density of possible emission sources, including transportation sources.

10.1.3 Datasets with Chemical Information

CHEMEST VALIDATION

DATASET NAME	STATUS	TITLE
CHCHEM	ON-LINE	CHEMICAL INFORMATION DATA FOR CHEMEST DEVELOPMENT
CHREF	ON-LINE	CHEMICAL REFERENCE DATA FOR CHEMEST DEVELOPMENT
CHBOILPT	ON-LINE	CHEMICAL BOILING POINT DATA FOR CHEMEST DEVELOPMENT
CHLOGP	ON-LINE	CHEMICAL LOG P DATA FOR CHEMEST DEVELOPMENT
CHWATER	ON-LINE	CHEMICAL WATER SOLUBILITY DATA FOR CHEMEST DEVELOPMENT
CHPKA	ON-LINE	CHEMICAL pKa DATA FOR CHEMEST DEVELOPMENT
CHVAPOR	ON-LINE	CHEMICAL VAPOR PRESSURE DATA FOR CHEMEST DEVELOPMENT

Compiled by the EPA Exposure Evaluation Division, Office of Toxic Substances, the CHEMEST Validation dataset includes many of the well-characterized chemical substances which are readily available in published literature.

Seven data files were individually implemented as GEMS datasets, including a chemical file (CHCHEM), a reference file (CHREF), and five data files, one for each of the following physiochemical properties: water solubility (CHWATER), octanol/water partition coefficient (Log P) (CHLOGP), boiling point (CHBOILPT), pKa (CHPKA), and vapor pressure (CHVAPOR).

CHCHEM contains information about the chemicals themselves, such as names and synonyms, Chemical Abstracts Service (CAS) registry numbers, molecular

formulas, and codes for the presence of chemical functional groups. These codes are a subset of 13 of the functional group codes found in OTS Technical Data Indexing System (TDIS) and were determined to be present in over 70 percent of nonpolymeric Premanufacture Notices (PMNs) stored in TDIS in May, 1983. The chemical file also contains an arbitrarily assigned chemical identification (CID) number which can be used as a key to access data in other files.

CHREF contains a record for each bibliographic source from which data have been extracted and input to one of the five property data files. Each record also contains a unique reference number and a flag variable to indicate whether it is a primary or secondary reference. The reference number provides an identifier used in the property data files to link the bibliographic data.

The property data files, CHBOILPT, CHLOGP, CHVAPOR, CHPKA, and CHWATER contain both measured data taken from the literature and estimated data from CHEMEST. Each record in the five property data files contains a CID and CAS number, a property value, its units, and the values and units of the most relevant experimental conditions such as temperature and pressure. If you want to estimate data values, the records also include an estimation method name, number, equation number and values of selected input data used in the calculations. The files contain multiple values for both measured and estimated data. An accession number is assigned to each record for a particular property for a specific chemical to segregate these multiple values. This accession number, combined with a CID or CAS number, provides a unique retrieval key for each record.

10.1.4 Geographic and Cartographic Datasets

REACH FILE

DATASET NAME	STATUS	TITLE
REACH	ON-LINE	STREAM INFORMATION

The REACH file was created by EPA's Monitoring and Data Support Division, Office of Water Regulations and Standards.

It contains information on more than 72,000 segments associated with more than 2,000 USGS cataloging units.

All stream reaches have been linked hydrologically to allow for retrieval and analysis downstream or upstream of a point or between points. The term "reach" is defined as a portion of a river which extends downstream from the confluence of two rivers (or from the uppermost end of a river) to the next encountered confluences or any subdivision of such a portion of a river. The reach is identified by an 11 digit number which combines the USGS HU code (8 digits) and EPA's segment number (3 digits). The REACH file in GEMS contains stream reach information for more than 68,000 reaches throughout the U.S. Primary variables for each reach in the dataset include:

- Cataloging unit code
- Segment number
- Mile point
- Latitude/longitude for different sections of the stream
- State FIPS code
- County FIPS code

You will find the REACH file indispensable for hydrologic modeling. You will also find that the dataset is associated with three other files that are linked together by common use of the variables: cataloging units, segment numbers, and mileages. These files are:

- Industrial Facilities Discharge (IFD)
- Water Supply Database (WSDB)
- Stream Gaging Inventory Data (GAGE)

SIC2EST

DATASET NAME	STATUS	TITLE
SIC2EST	ON-LINE	TWO-DIGIT SIC ESTABLISHMENTS LISTING

The SIC2EST file was created using data extracted from the 1983 County Business Patterns Survey done by the Bureau of the Census, to provide a listing of the number of business establishments in an area by two-digit SIC code.

The file contains the state and county FIPS codes, along with the first two digits of the SIC codes and number of the business establishments in the identified county.

This data can be useful in a number of GEMS modeling applications.

SIC4CNTY

DATASET NAME	STATUS	TITLE
SIC4CNTY	ON-LINE	INDUSTRIES BY 4 DIGIT SIC CODES BY COUNTY

The SIC4CNTY file consists of information extracted from the 1983 County Business Patterns Survey specifically for use with GEMS models.

The dataset includes state and county FIPS codes, the 4 digit SIC code and the numbers of establishments and employees in those establishments.

Use of this data can be helpful in numerous GEMS modeling activities.

ZIPCODE

DATASET NAME	STATUS	TITLE
ZIPCODE	ON-LINE	ZIP CODE ASSOCIATED WITH LATITUDE AND LONGITUDE IN DEGREES AND MINUTES

The ZIPCODE file was created as a proprietary product of Geographic Systems and was obtained by EPA Region II. The file associates U.S.P.S. zip codes with the appropriate latitude and longitude coordinates of the zip code area centroids or with post offices servicing each area. This file allows you to

obtain geographic coordinates based on zip code, when latitude/longitude coordinates are not otherwise available.

Release site coordinates are needed in locating data from nearby weather and other monitoring stations and in identification of populations living nearby. The ZIPCODE file is useful when release site latitude and longitude coordinates are not available but are needed in an exposure analysis.

FIPSCODE

DATASET NAME	STATUS	TITLE
FIPSCODE	ON-LINE	FIPS CODES OF US COUNTIES EXCEPT FOR ALASKA

The FIPSCODE dataset contains the five-digit state and county Federal Information Processing System (FIPS) codes, the state name, and the county name. It can be used to look up the name of a county when the user has the appropriate FIPS code for that county.

USCITIES

DATASET NAME	STATUS	TITLE
USCITIES	ON-LINE	US CITY NAMES AND THEIR LOCATIONS

The USCITIES file was created from the data used for the ZIPCODE file. Only those ZIPCODES which have a city name associated with it were retained. The file includes a rough estimate of the cities latitude/longitude which can be used when no other location data are available.

PLACES

DATASET NAME	STATUS	TITLE
PLACES	ON-LINE	ALL POPULATED PLACES IN THE US

General Sciences Corporation combined information from several different sources to produce a dataset that helps identify virtually all populated places in the United States. The data includes 160,000+ records showing the FIPS names, location in latitude/longitude coordinates and geographic size, along with the identity of the larger geographic entities in which the places are located. For instance, Silver Spring, Maryland would include the notation that it was in Montgomery County.

10.2 Auxiliary Files

These files are used by the modeling routines to accomplish their objectives. They are not accessible separately as is the case with the files discussed in section 10.1. These files are discussed in this section so that you will understand how they work.

10.2.1 Datasets to Support Environmental Models

1980 MASTER AREA REFERENCE FILE (MARF)

CEDPOP is a subset of the MARF of the 1980 Census. In addition to total population and household counts, the file includes geographic coordinates for the population-weighted centroid of each census block group or enumeration district (BG/ED) in the file.

The CEDPOP file is accessed by CENSUS procedure under the GEMS Geodata Handling operation. The CENSUS procedure accumulates population and housing counts by up to ten user-specified radial distances and from one to sixteen sectors.

NATIONAL SOILS DATABASE

This information on soil characteristics was compiled for the 48 contiguous states by Battelle Northwest Laboratories under contract to EPA, and is an indexed, direct-access file for which the index is keyed to any latitude and longitude in the continental U.S. Available data include groundwater region,

state FIPS code, soil type, erosion, top soil field capacity, distance to nearest drinking water well, high and low values for depth to the water table, effective porosity, and groundwater flow.

This database is accessed by GEMS models including SESOIL, TOXSCREEN, AT123D, whenever effective porosity, depth to groundwater, and other soil information are needed.

CANONICAL ENVIRONMENTS DATABASE

Battelle Northwest Laboratories under contract to the EPA laboratory in Athens, Georgia, developed sets of input file parameters for use with the EXAMS-II model. This database has been expanded by the Exposure Evaluation Division of EPA's Office of Toxic Substances.

Two data files have been developed for the definition of aquatic canonical environments:

- Riverine data file
- Lakes/reservoirs data file

The two files are necessary because the temporal and spatial structure of lakes and reservoirs differ from riverine systems and, therefore, require different sets of variables for their description. Together, these two files embody the Canonical Environments Data Base (CEDB).

The database contains sets of most of the EXAMS-II model input parameters for a number of major U.S. river systems (85 rivers); and more than 700 lakes and reservoirs. Included in the file are parameters such as:

- River reach flow
- Suspended sediment concentration with organic carbon content
- Temperature
- Wind velocity

- Rainfall
- Surface runoff
- Sediment loadings
- Groundwater interflow

TABLES

Rather than provide tables of common codes in each chapter, these data are presented in this section for quick reference by readers of individual chapters. Codes used by various functions have been combined, wherever possible, to reduce the amount of space required to present the information; therefore, these tables may include columns providing more information than is required by individual functions.

Where a table is unique to a single function, that table may be included within the appropriate chapter. Inclusion in this section is intended only to reduce the amount of space required to support redundant pages.

Tables in this section include:

1 State Names

- An alphabetical listing of state names with associated USPS codes and FIPS codes.

2 Phone Numbers

- Important phone numbers for GEMS access and support.

TABLE 1 - State Codes

STATE NAME	USPS	FIPS
ALABAMA	AL	01
ALASKA	AK	02
ARIZONA	AZ	04
ARKANSAS	AR	05
CALIFORNIA	CA	06
COLORADO	CO	08
CONNECTICUT	CT	09
DELAWARE	DE	10
D.C.	DC	11
FLORIDA	FL	12
GEORGIA	GA	13
HAWAII	HI	15
IDAHO	ID	16
ILLINOIS	IL	17
INDIANA	IN	18
IOWA	IO	19
KANSAS	KS	20
KENTUCKY	KY	21
LOUISIANA	LA	22
MAINE	ME	23
MARYLAND	MD	24
MASSACHUSETTS	MA	25
MICHIGAN	MI	26
MINNESOTA	MN	27
MISSOURI	MO	29
MONTANA	MT	30
NEBRASKA	NE	31
NEVADA	NV	32
NEW HAMPSHIRE	NH	33
NEW JERSEY	NJ	34
NEW MEXICO	NM	35

TABLE 1 - State Codes (Cont'd)

STATE NAME	USPS	FIPS
NEW YORK	NY	36
NORTH CAROLINA	NC	37
NORTH DAKOTA	ND	38
OHIO	OH	39
OKLAHOMA	OK	40
OREGON	OR	41
PENNSYLVANIA	PA	42
RHODE ISLAND	RI	44
SOUTH CAROLINA	SC	45
SOUTH DAKOTA	SD	46
TENNESSEE	TN	47
TEXAS	TX	48
UTAH	UT	49
VERMONT	VT	50
VIRGINIA	VA	51
WASHINGTON	WA	53
WEST VIRGINIA	WV	54
WISCONSIN	WI	55
WYOMING	WY	56

TABLE 2 - Phone Numbers

	FTS	OTHER
VAX USER SUPPORT	629-7862	(919)541-7862
		(800)334-2405
RTP STATUS PHONE	629-2969	(919)541-2969
RTP TELECOMMUNICATIONS SUPPORT SERVICE	629-4506	(919)541-4506
TYMNET USER SUPPORT (Voice)		(800)336-0149

REFERENCES

Bonazountas, M. and Wagner, J.M., 1984. SESOIL: A Seasonal Soil Compartment Model. Cambridge, MA: Arthur D. Little, Inc. for Office of Toxic Substances, U.S. EPA. NTIS Order No. PB86-112406

Burns, L.A. and Cline, D.M., 1982. Exposure Analysis Modeling System (EXAMS): User Manual and System Documentation. Athens, GA: Environmental Research Laboratory, U.S. EPA. NTIS Order No. PB82- 258096

Burns, L.A. and Cline, D.M., 1985. Training Guide and Reference Manual for EXAMS II. Athens, GA: Environmental Research Laboratory. PB85-228138/LP

Carsel, R.F., Smith, C.N., Mulkey, L.A., Dean, D.J., and Jowiss, P., 1984. User's Manual for the Pesticide Root Zone Model (PRZM) - Release 1. Athens, GA: Environmental Research Laboratory, U.S. EPA. NTIS Order No. PN85-158913/LT

Census, 1983. 1980 Census of Population. Volume 1 Characteristics of Population. Number of Inhabitants. PC80-1-A1.

Donnelly Marketing Information Services, 1983. Annual Demographic Estimates.

Hanna, S.R., 1964. Dry Deposition and Precipitation Scavenging in the ATDL Computer Model From Multiple Point and Area Sources. Oak Ridge, TN: Air Resources Atmospheric Turbulence and Diffusion Laboratory, ATDL Contribution File No. 71.

Hanna, S.R., 1977. A Stability Correction Term for a Simple Urban Dispersion Model. Oak Ridge, TN: Air Resources Atmospheric Turbulence and Diffusion Laboratory, ATDL Contribution File No. 77/15.

Hanna, S.R., 1980. Atmospheric Removal Processes for Toxic Chemicals. Draft. Oak Ridge, TN: ATDL Progress report to ORNL Under EPA Multimedia Modeling Project.

Hetrick, D.M. and McDowell-Boyer, L.M., 1984. User's Manual for TOX-SCREEN: A Multi-Media Screening Level Program for Assessing the Potential Fate of Chemicals Released to the Environment. Oak Ridge, TN: Oak Ridge National Laboratory for the U.S. Department of Energy (DOE).

Office of Air Quality Planning and Standards, 1986. Guideline on Air Quality Models (Revised). Research Triangle Park, NC: Office of Air Quality Planning and Standards, U.S. EPA. NTIS Order No. PB88-150958

Petersen, W.B., Catalano, J.A., Thomas, C., and Yuen, T.S., 1979. INPUFF - A Single Source Gaussian PUFF Dispersion Algorithm User's Guide. Research Triangle Park, NC: Environmental Sciences Research Laboratory, U.S. EPA. NTIS Order No. PB85-137131

Pierce, T.E. and Turner, D.B., 1982. PTPLU - A Single Source Gaussian Dispersion Algorithm. Research Triangle Park, NC: Environmental Sciences Laboratory, U.S. EPA. NTIS Order No. PB83- 211235

Turner, D.B. and Susse, A.D., 1973. User's Guide to the Interactive Versions of Three Point Source Dispersion Programs: PTMAX, PTDIS, and PTMTP. Research Triangle Park, NC: National Environmental Research Center, U.S. EPA.

General Sciences Corporation, MD, 1985. User's Guide to SWIP Model Execution Using Data Management Support System. Laurel, MD: General Sciences Corporation. GSC-TR8531

General Sciences Corporation, 1985. User's Guide to Environmental Partitioning Model. Laurel, MD: General Sciences Corporation. GSC-TR8518

General Sciences Corporation, 1985. User's Guide to AT123D Execution. Laurel, MD: General Sciences Corporation. GSC-TR8542

General Sciences Corporation, 1985. User's Guide to Microbe-Screen Execution in GEMS. Laurel, MD: General Sciences Corporation. GSC-TR8753

General Sciences Corporation, 1985. Characterization of Data Base Requirements for Implementation of UTM-TOX under GEMS: Parameter Sensitivity Study. Laurel, MD: General Sciences Corporation.

General Sciences Corporation, 1986. Groundwater Modeling Scenarios for Assessment Applications: Part II - Groundwater Dispersion Information from Case Studies. Laurel, MD: General Sciences Corporation. GSC- TR8641.

General Sciences Corporation, 1986. INPUFF User's Guide. Laurel, MD: General Sciences Corporation. GSCTR8621

General Sciences Corporation, MD, 1987. User's Guide to SESOIL Execution in GEMS. Laurel, MD: General Sciences Corporation. GSC. GSC-TR8747

General Sciences Corporation, 1987. User's Guide to EXAMS II Execution in GEMS. Laurel, MD: General Sciences Corporation. GSC-TR8710

General Sciences Corporation, 1988. User's Guide for PTPLU in GEMS. Laurel, MD: General Sciences Corporation. GSC-TR-32-88-019

General Sciences Corporation, 1988. PCGEMS User's Guide. Laurel, MD: General Sciences Corporation. GSC-TR-32-88-027

General Sciences Corporation, 1987. Implementation of the Climatic Data Base in GEMS. Laurel, MD: General Sciences Corporation. GSC-TR8755

General Sciences Corporation, 1987. User's Guide to TOX-SCREEN Execution in GEMS. Laurel, MD: General Sciences Corporation. GSC-TR8751

General Sciences Corporation, 1989. GEMS Atmospheric Modeling System (GAMS) User's Guide. Laurel, MD: General Sciences Corporation. GSCTR-32-88-017

Patterson, M.R., Sworski, T.J., Sjoreen, A.L., Browman, M.G., Coutant, C.C., Hetrick, D.M., Murphy, B.D., Raridon, R.J., 1984. A User's Manual for UTM-TOX, the Unified Transport Model. Oak Ridge, TN: Oak Ridge National Laboratory. NTIS Order No. PB- 213750/LT

Yeh G.T., 1981. AT123D: Analytical Transient One-, Two-, and Three-Dimensional Simulation of Waste Transport in the Aquifer System. Oak Ridge, TN: Oak Ridge National Laboratory. NTIS Order No. ORNL-5602/LT

Wackter, D.J. and Foster, J.A., 1986. Industrial Source Complex (ISC) Dispersion Model User's Guide - Second Edition. East Hartford, CT: TRC Environmental Consultants, Inc. for the Office of Air Quality Planning and Standards, U.S. EPA. NTIS Order No. PB86-234259/LP.

INDEX

A

Access Census Data Procedure (CENSUS), 4-10

Archived files, 1-15, 1-22

AT123D, 1-23

AUTOCHEM, 1-4, 1-31

AUTOHELP command, 1-31

B

BACK command, 1-31

BGEDMAP procedure, 4-42

Block Groups and Enumeration District (BG/ED), 1-3

BOXMOD, 1-16

C

Canonical Environments data, 1-29

CAS, 1-6

CEDPOP subset, 1-28

Census database selections, 4-7

CHEMEST, 1-3, 1-16

- Acid Dissociation Constant, 1-30
- Activity Coefficients, 1-23
- Bioconcentration Factors for Fish, 1-22
- Boiling Point, 1-24
- Henry's Law Constant, 1-28
- Melting Point, 1-29
- Rate of Volatilization from Water, 1-27
- Soil Adsorption Coefficient, 1-22
- Solubility in Water, 1-21
- Vapor Pressure, 1-26

CHEMEST Commands

- ?, 1-19
- CHEM, 1-20
- CMD, 1-20
- DISPLAY, 1-20

END, 1-19

FILE, 1-20

FLT, 1-20

FORMULA, 1-20

HELP, 1-19

INT, 1-20

LAST, 1-20

METH, 1-20

NAM, 1-20

NEW, 1-20

OLD, 1-20

PMT, 1-20

PROP, 1-20

RESTART, 1-19

STOP, 1-19

SUM, 1-20

CHEMEST Validation dataset, 1-22

Chemical Abstacts Service (CAS) number, 1-23

Chemical Property Estimation, 1-23

CLEAR command, 1-31

Climate database, 1-5

CLOGP3, 1-3, 1-11

UDRIVE Access System, 1-12

Using UDRIVE, 1-13

CNTYSIC dataset, 1-19

Commands, 1-30

Communication Methods to GEMS, 1-35

Communication with PCs, 1-33

Contigency Tables procedure, 1-1

Convert Geocode to Lat/Long (LATLON) procedure, 4-27

CrossTalk, 1-37

D

Data Transformation Procedure, 1-14

Dataset Definition, 1-1

Datasets, brief definition of, 1-25

DECnet access, 1-6

Descriptive Statistics procedure, 1-3

Determine County Coverage (COVERAGE) procedure, 4-16

Disperse dataset, 1-7

DRASTIC dataset, 1-8

Draw County Map (CNTYMAP) procedure, 4-44

DRAWSMI, 1-3, 1-10

E

Economic Census data, 1-21

ENPART, 1-35

Equations in tabulation variables, 4-22

EXAMS II, 1-31

Canonical Environments Data Base (CEDB), 1-32

EXIT command, 1-31

Export Files

Convert a GEMS File to VAX Format, 1-28

Extract Procedure, 1-15

F

FAP, 1-38

FAP Commands

CALC, 1-42

DIRS, 1-42

DONE, 1-42

DRAW, 1-42

HELP, 1-42

HETR, 1-43

LIST, 1-43

NOPU, 1-43

PAUS, 1-43

PRIN, 1-43

TEKT, 1-43

TERM, 1-43

WKLS, 1-43

File Management

Change Status Option, 1-35

Create Datasets Option, 1-3

Delete Option, 1-23

Import Option, 1-31

Modify Datasets Option, 1-18

Export Option, 1-27

Output Option, 1-24

Find US Cities procedure (USCITY), 4-34

FIPSCODE dataset, 1-27

FRDSPWS dataset, 1-11

G

GAGE database, 1-12

GAMS, 1-3

Exposure and risk calculations, 1-4

Reports, 1-10

GEMS Components, 1-2

Geodata Handling Mapping (GEOMAP) procedures, 4-41

Geoecology database, 1-13

Geographic Data Lookup and Display (GEOLIST), 4-2

Geographic Data Management (GEODM) procedure, 4-19

GEOLIST, 4-1

GEOMAP, 4-1

Graph Definition, 1-1

GRNDWAT dataset, 1-8

H

HELP command, 1-32

How GEMS Works, 1-1

HUCODE/SOIL locator (HUCODE) procedure, 4-23

HWDMSFacility file, 1-19

Hydrological Unit Map (HUMAP) procedure, 4-56

I

IFDDIR, 1-16

IFDIND, 1-16

Industrial Facilities Discharge (IFD) file, 1-16

INPUFF, 1-11

Internation access via TYMNET, 1-8

Introduction, 1-1

ISCLT, 1-6

K

Kermit, 1-35

L

Latitude/Longitude Locator System, 4-2

LIST command, 1-32

Log-In, 1-7

Lookup/Examine STAR Station Data, 4-31

M

Map Definition, 1-1

Map Unit Use File (MUUF), 1-5

Master Area Index File (MARF), 1-2

MENU command, 1-32

MICROBE-SCREEN, 1-40

Multiple Regression procedure, 1-8

N

National Soils Database, 1-28

NEEDS86 file, 1-18

NEWCHEM, 1-4, 1-30

NEXT command, 1-32

NO AUTOHELP command, 1-32

P

Password, 1-10

PLACES dataset, 1-27

Port Selector and Access, 1-4

Port Selector at RTP, 1-9

POTW, 1-16

Project, 1-11

PROMPT command, 1-32

PRZM, 1-22

PTPLU, 1-13

Q

QUEUE command, 1-32

R

REACH file, 1-24

Regional access numbers table, 1-6

Regional dial access, 1-5

RESET command, 1-32

RESTORE command 1-33

REVIEW command, 1-33

S

Save Temporary Dataset procedure, 1-36

SESOIL, 1-18

SFILES, 1-1, 1-3, 1-5

SHOW command, 1-33

SIC2EST dataset, 1-25

SIC4CNTY file, 1-26

Simple Regression procedure, 1-7

Site Level Retrieval Of Data (SITERET), 4-4

SMIGET, 1-3, 1-6

SMILES, 1-1 - 1-2

SMILES Tips, 1-4

SMILES Notation, 1-2

Nomenclature Rules, 1-2

SOILS-5 database, 1-5

STAR dataset, 1-4

STATESIC dataset, 1-18

Statistics Operation, 1-1

STORE command, 1-33

SWIP, 1-28

T

Table Definition, 1-1

TANKSITES

Primary Sampling Units (PSU) regions, 1-10

TANKSITES dataset, 1-9

TGEMS, 1-30

The State/County Map Procedure (STCOMAP), 4-62

TOX-SCREEN, 1-37

TOXBOX, 1-7

TYMNET Customer Support and Trouble-Shooting, 1-9

U

Universal Transverse Mercator (UTM) system, 4-2

USCITIES file, 1-27

USERCMD command, 1-33

UTM-TOX, 1-41

V

VAX Cluster and GEMS, 1-4

VT100 Keypad Definitions, 1-8

VULNERGW dataset, 1-8

W

Washington D.C. Access, 1-5

Who uses GEMS, 1-1

WSDB database, 1-13

Z

Zip Code Geographic Location Method, 4-2

ZIPCODE file, 1-26

GEMS User's Response Form

We appreciate your feedback on this User's Guide. If you could take a few moments to share your comments, it will help us in the continuing effort to provide you with professional user documentation. Please share your evaluation of this document's:

Content

Organization

Readability

Style & Format

Specific Problems, or Comments:

Name:

Agency:

Address:

City, State, ZIP:

Daytime Phone:

PleaseMail to:
GEMS Support Group
General Sciences Corporation
6100 Chevy Chase Drive
Laurel, MD 20707
or: FAX (301)953-2941